DOCTOR OF PHILOSOPHY

The Use of Social Gaze Cues in Real World Scenes

Mitchell, Kathryn Mary Anne

Award date:
2015

The Use of Social Gaze Cues in Real World Scenes

Kathryn Mary Anne Mitchell

MA (Hons.)

University of Dundee

January 2015

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of Philosophy in Psychology

Table of Contents

List of Figures

Acknowledgements

Much like any great endeavour, a thesis is not completed alone. It takes a community to ensure the first flurry of enthusiastic but under-developed ideas are nurtured into a sizeable piece of work that represents not only a great deal of effort and a passion for the topic, but a meaningful contribution to the chosen field. I have had the benefit of a sizeable community who have supported me throughout my PhD tenure, and to them I wish to give thanks.

Of course, I must first thank the School of Psychology for funding my PhD studentship, because without that backing I could never have aspired to even begin my research. It was my supervisor, Ben, who supported that application and without him I wouldn’t have made it to submission, so thank you Ben for all the hard work you have put in over the last three-and-a-bit years, and for helping me to shape this work into something I can be proud of. It has been my privilege.

I have met many wonderful people throughout my time as a postgraduate at the University of Dundee, staff and students alike, and it has been great to be a part of your community. Some of you have already gone on to new adventures, and I hope our paths cross again. To Wayne, Astrid, Heather, Fhionna, Glenn, Ross, Katja, Laura and Juliet, thank you. My most special thanks to Anne and to Jaime, who do not get to stop being pillars of my own community yet! I have also had the pleasure of working with four wonderful undergraduates, and I wish you all the very best with your future endeavours, wherever they take you – I know you will do very well. Gemma, Anne, Julia, and Nicole, thank you. Particular thanks to Nicole and Julia who allowed their images to be used in the multiple gaze cue study stimuli.

My supportive community has many members from out with the university, too. I would be entirely neglectful if I didn’t express my gratitude to my family who have cheered me on throughout, not just my mum and dad, but my ‘Cupar tribe’ as well. You know how much it has meant to me. I am also lucky enough to have several friends who are full of wisdom and words of advice, who have always believed in me, and whose belief has kept me going: Emily, Yvonne, Susan and Hannah, thank you.

My final thanks are dedicated to my husband, Will, who generously volunteered to be the subject in my single cue stimuli. I know that living with a PhD student writing up a thesis is an endeavour all of its own, and your endless patience and unwavering faith in me have earned many more brownie points than I can hope to repay! I am so glad to be able to share this achievement with you, and while it may not be a sonnet, a song, or a great work of art, this work that I have made is dedicated to you.

Dissemination of Research

Publications

Mitchell, K. M. A., & Tatler, B. W. (in prep). An extra pair of eyes: multiple gaze cues in social scenes.

Tatler, B. W., Kirtley, C., Macdonald, R. G., **Mitchell, K. M. A.,** & Savage, S. W. (2014). The active eye: perspectives of eye movement research. In M. Horsley, M. Eliot, B. Knight, & R. Reilly (Eds.). *Current Trends in Eye Tracking Research.* London: Springer International Publishing.

Conference Presentations

Mitchell, K. M. A., & Tatler, B. W. (2014). Multiple gaze cues and their effect on observer eye movements in a visual search task. Talk presented at the Scottish Vision Group (SVG) Conference, Troon, UK, 22nd of March 2014.

Mitchell, K. M. A., & Tatler, B. W. (2013). Following Multiple Gaze Cues in Social Scenes. *Talk presented in the Special Symposium on Social Gaze at the 17th European Conference on Eye Movements (ECEM), University of Lund, Lund, Sweden, 14th of August 2013.*

Mitchell, K. M. A., & Tatler, B. W. (2013). How multiple gaze cues affect our viewing of social scenes. *Talk presented at the Fife-Tay Vision Group meeting, University of Abertay, Dundee, UK, 25th of July 2013.*

Declaration

The candidate is the author of this thesis. Unless otherwise stated, all references cited have been consulted by the candidate. The work of which the thesis is a record has been done by the candidate, and has not been previously accepted for a higher degree.

Signed: ______________________________

(Kathryn M. A. Mitchell)

Date: ______________

Thesis Abstract

Eyes are an ideal tool for investigating social attention, as their physiological composition with the iris and pupil highly-distinguishable against the white sclera, combined with our foveated vision, mean that gaze cues are both a means of understanding where attention is being allocated and a method for non-verbal communication. Previous attention research using gaze cues has focused on Posner-type paradigms that have supported a model of reflexive orienting of attention in response to gaze cues. However, the ecological validity of this type of paradigm has been called into question given more recent real world research, which has produced findings that cannot be explained by laboratory-based Posner-type paradigms. Therefore, the aim of this thesis was to develop and test a novel, more ecologically-valid paradigm that could investigate observers' responses to gaze cues in a realistic, but controlled, manner.

Based on past research, an initial goal of this research was to develop an early iteration of a realistic visual search paradigm in which a single non-predictive gaze cue is presented. This was built on in later chapters by adding manipulations of task instruction. These chapters presented some evidence that supported a reflexive orienting model of gaze, with clear facilitation to performance as a result of person presence. The second goal of this research was to explore observers' responses when presented with the same task and search arrays, but with the inclusion of a second gaze cue. This is some of the first research to address multiple gaze cues within a realistic visual search paradigm. These chapters showed multiple gaze cues result in quite considerably different observer eye movement behaviour. Benefits of

people presence were stronger and far more congruency effects were apparent. There were also clear effects of instruction, with the suggestion that gaze cues provided may be helpful to the task resulting in significantly greater proportions of overt gaze-seeking than in other instruction conditions. The introduction of multiple gaze cues created a new gaze cue condition – the *conflicting* condition in which each person cued separate spatial areas within the scene. In order to explore the effects of gaze cue sender reliability on observers' eye movements, a third version of the study was tested where the gaze cues presented were spatially informative, cuing the target in 70% of trials. Results showed similar benefits of people presence to the previous multiple-cue chapters, but there were minimal reliability effects. Methodological adaptations were suggested based on previous research that has explored reliability effects that may more successfully elicit reliability effects in future research.

The final chapter presents a summary of the findings of the research contained within this thesis. The results showed that in a more complex and realistic visual search task employing a single gaze cue, results are somewhat consistent with the reflexive orienting model of gaze due to the clear facilitation as a result of person presence and the lack of instruction effects. The findings presented also demonstrate that once multiple gaze cues are introduced, the reflexive orienting model cannot account for observers' gaze behaviour. Instead, findings are more consistent with recent real world research. This would suggest that a new model of gaze processing is required when more than one gaze cue is presented, and the final chapter offers some suggestions of what this new model would need to take into account. It is suggested that subsequent research using this novel paradigm should explore the use of dynamic cues and the effects on eye movement behaviour in special populations, and that the research presented in this thesis provides a solid foundation upon which these new directions for research can be built.

Chapter One

Investigations of Social Attention

The Importance of Eyes

The evolution of humans to live in social groups has ensured our survival. Living in groups makes it easier to divide the workload of gathering resources, to share these resources, and to share the responsibilities of looking out for danger or caring for the very young, elderly or infirm (Frith, 2008). Our brains and behaviours have adapted to make each of these shared interactions easier, allowing us to use social behaviours to communicate and learn from each other (Frith & Frith, 2007). Eyes are perhaps the most critical tool for human social interaction, and it is their structure that makes them ideal for non-linguistic communication. Kobayashi and Koshima (1997) identify the anatomy of the human eye as unique. With the coloured iris and dark pupil easily distinguishable from the white sclera, the gaze direction of people around us can be very quickly detected. Non-human primates, such as orang-utans and chimpanzees do not have this highly visible sclera; their sclera is similar in colour to the skin around their eyes, and therefore gaze direction detection is more difficult.

Humans, amongst many other animals including fish, crabs and cuttlefish, exhibit a pattern of eye movements known as ‘saccade and fixate’ (Land, 1999). Humans fixate on something in the environment by holding the eye still for an average duration of 300ms before launching a saccade – the fast movement of the eye between fixations. Humans have foveated vision, which means that only the centre of our visual field is seen in high detail. Therefore, to examine something closely, this stimulus needs to be centred on the fovea so that it can be seen at the highest level of detail (Land & Tatler, 2009). In essence, this means that when we want to look at something we point our eyes at it. Having an easily identifiable ‘pointer’ that is directed to whatever we are attending provided our early ancestors with a method of understanding what another person was thinking or intending before they had the power to communicate this verbally. It is this physiology which allows humans to detect even very small changes in eye gaze direction, which can prove invaluable when gathering resources or defending against a predator.

The ability to use these kinds of physical cues has been tested against the cognitive skills of our nearest primate relatives: chimpanzees and orang-utans. Herrman, Call, Hernández-Lloreda, Hare and Tomasello (2007) tested large groups of these primates in their cognitive skills in comparison to human infants up to the age of two-and-a-half years who as yet were not fully capable of expressing themselves via verbal language and had not been influenced by written language or formal education. Herrman et al. (2007) found that while human infants and primates perform relatively similarly on tasks involving tool manipulation, human infants were better at causality tasks and those which involved early processes of Theory of Mind. This ability to understand the mental states of others, which is discussed in more detail later in this chapter, is one of the key cognitive skills that distinguish us from our primate relatives. Baron-Cohen, Wheelwright and Jolliffe (1997) call this the ‘language of the eyes’, through which humans can detect emotions from seeing the eyes only. Within the context of a whole face, we are able to detect very subtle or complex emotional shifts. This study was the first to demonstrate eyes alone contain enough information to detect complex mental states; a skill which would be invaluable to our early ancestors who communicated with little or no verbal language.

Defining Social Attention

Development of social attention in infancy

This type of non-verbal communication falls under the umbrella term of social attention, which describes a number of closely-related cognitive processes including social referencing, joint attention, Theory of Mind and gaze seeking and following. The production of these behaviours and processes are milestones in infant development (Carpenter, Nagell, & Tomasello, 1998) because they demonstrate that a child is learning that they are a unique self with different thoughts, ideas and knowledge to others. Gaze seeking and following behaviours, which are part of joint attention, are valuable for the learning infant. By first seeking and then following the caregiver's gaze the infant can identify new objects to interact with. Following the gaze of the caregiver results in jointly attending the object with them, and the infant can use the caregiver's other cues – such as facial expression – to learn about the object. Thoermer and Sodian (2001) report that infants as young as 12-months-old are found to respond to objects cued by adults' gaze. Confronted with something new that they are unsure of, an infant will look to their caregiver. Their reaction will inform the infant of the appropriate response – a frown suggests avoidance, a smile encourages approach. This is called social referencing, which Feinman (1982) defined as an emotional communication with someone – usually a caregiver – where the infant uses their understanding of the caregiver's interpretation of an event to process the situation. Gergely, Egyed and Kiraly (2007) found 14-month-old infants could use social referencing to learn about the value of objects. By 18-months-old, they could begin to understand whether another person liked the object or not. This sort of learning has been well documented and demonstrated in a reinterpretation of the classical Gibson and Walk (1960) 'Visual Cliff', which also demonstrates infants' use of social referencing. The original study, designed to test depth perception in infants, simulated a visual drop by having a clear Plexiglas sheet above a small drop. This made the drop visible, but infants were in no danger. Sorce, Emde, Campos and Klinnert (1985) revisited the paradigm with mothers displaying a variety of emotions when their infant child looked to them for guidance. The results showed clearly that infants used their mother's expressions to disambiguate the situation and to regulate behaviour; joy or interest would encourage them to cross, whereas fear or anger resulted in most infants staying where they were.

Theory of Mind is more complex and often understood from a goal-driven perspective. Social goals involving at least two people require collaboration between both participants, and their goals can be altruistic, competitive or co-operative (Frith, 2008). Theory of Mind usually develops between the ages of three and four years in human infants, and is the means by which a child is able to predict or explain another's behaviour by that person's feelings, thoughts or beliefs, which may be different from their own (Wilmmer & Perner, 1983). Traditionally this is tested for using a false belief paradigm. A well-known example of this is the Sally Anne task, developed by Baron-Cohen, Leslie and Frith (1985). In this task, a child is shown two puppets, Sally and Anne, who are playing together. Sally puts a marble in her basket then Anne leaves the room. While she is out of the room, Sally moves the marble from her basket into Anne's toy box. When Anne returns to the room the child is asked where she will look for the marble. It is only once the child has developed Theory of Mind and can understand that Anne will hold a belief that is wrong since she has not been witness to the moving of the marble, that they can correctly answer that Anne will look for the marble in Sally's basket.

Use of social attention in adulthood

While the essential components of social attention are developed during infancy, these behaviours are used in all social interactions throughout our lives. Despite our sophisticated language for verbal communication, we use gaze cues and other social signals to make swift judgements about those around us. Willis and Todorov (2006) demonstrated that we make decisions about other peoples' characteristics in as little as 100ms by showing participants limited exposure of peoples' faces. Across five experiments, they asked viewers to rate the person's trustworthiness, competence, attractiveness, likeability and aggressiveness. When exposure time was increased to 500ms, participants' ratings were likely to become more negative and they were more confident in these ratings. People use their understanding of faces and gaze cues to make assessments like this daily, and it has been found in many other studies that gaze cues are used to ascertain characteristics like trustworthiness (Bayliss & Tipper, 2006; King, Rowe & Leonards, 2011), as well as anger and aggressiveness (Ewbank, Jennings & Calder, 2009). The ability to make these kinds of judgements has been cited as having influence as far as our ability to sustain effective democracy via our understanding of others' political expertise (Huckfeldt, 2001).

What is additionally illustrative of our understanding of social attention is research documenting our understanding of how our own cues affect the perceptions of others. Chartrand and Bargh (1999) described the 'chameleon effect', which refers to the mimicry of postures, mannerisms and expressions of partners in interaction; an unintentional behaviour that makes us seem more likeable. Looking at someone's eyes allows understanding of what someone is feeling (Baron-Cohen, Baldwin & Crowson, 1997), to facilitate turn-taking in conversation and to establish a comfortable 'personal space' between two parties (Argyle & Dean, 1965). This sort of modulation of behaviour is particularly evident when people are engaged in, or are witness to, offensive behaviour. Crosby, Monin and Richardson (2008) examined how people respond to others' potentially offensive behaviour by tracking participant eye movements while watching a video discussion about university admissions between three white males and one black male. One of the white males criticised affirmative action, and participants were either led to believe the black discussant heard what was said via his headphones, or could not hear what was said. The results showed that participants would look to the black male when they believed he could hear the offensive comment, but not when they believed he had not heard it. Crosby et al. (2008) suggest this could be an example of social referencing, where participants are using the response of a potentially victimized individual to understand the situation and to appropriately inform their response.

Social Gaze and Attention Research

Previous investigation of attention

That the unique physiology of human eyes, discussed previously, creates a deictic ‘pointer’ cue used to guide behaviour (Shepherd, 2010), and the idea of eyes functioning as pointers to stimuli a person is interested in has been particularly important for the study of attention – the process by which we select one particular stimulus to inspect more closely (Buschman & Miller, 2010). Having the ability to limit the focus of attention is essential, as humans do not have the cognitive capacity to simultaneously process and hold in memory all the components of our environment. Visual search paradigms have been the main means by which attentional capture and stimulus selection are studied. Simons (2000) cites three main methodologies: the addition of a distinctive item appearing in the visual search array, which can be performed with or without eye movement recording; an irrelevant distinctive item appearing in the array at a target or a distractor location; or an irrelevant spatial cue given prior to the appearance of the search array.

One of the first models of attention was proposed by Treisman (1980), and was called the Feature-Integration Theory of attention. In this theory, it was suggested that when a visual search array is viewed without prior knowledge of its content, attention is allocated serially to each stimulus to find the unique target. It is important to note that Treisman (1980) was talking exclusively about attention shifts, and did not cite these shifts as dependent on eye movements. These early visual search paradigms used differentiation of geometric shapes, colour, alphanumeric characters or parts of figures, such as lines and curves. Figure 1 below shows a series of potential visual search tasks. In the top row, participants simply have to do a colour discrimination to find the red circle (present in panel B). In the middle row, they must perform shape discrimination in order to identify the red circle amongst red squares (panel D). The bottom row shows the most complex task, where observers must both perform colour discrimination to eliminate all blue targets, and then a shape discrimination to identify the red circle amongst red squares (panel E).

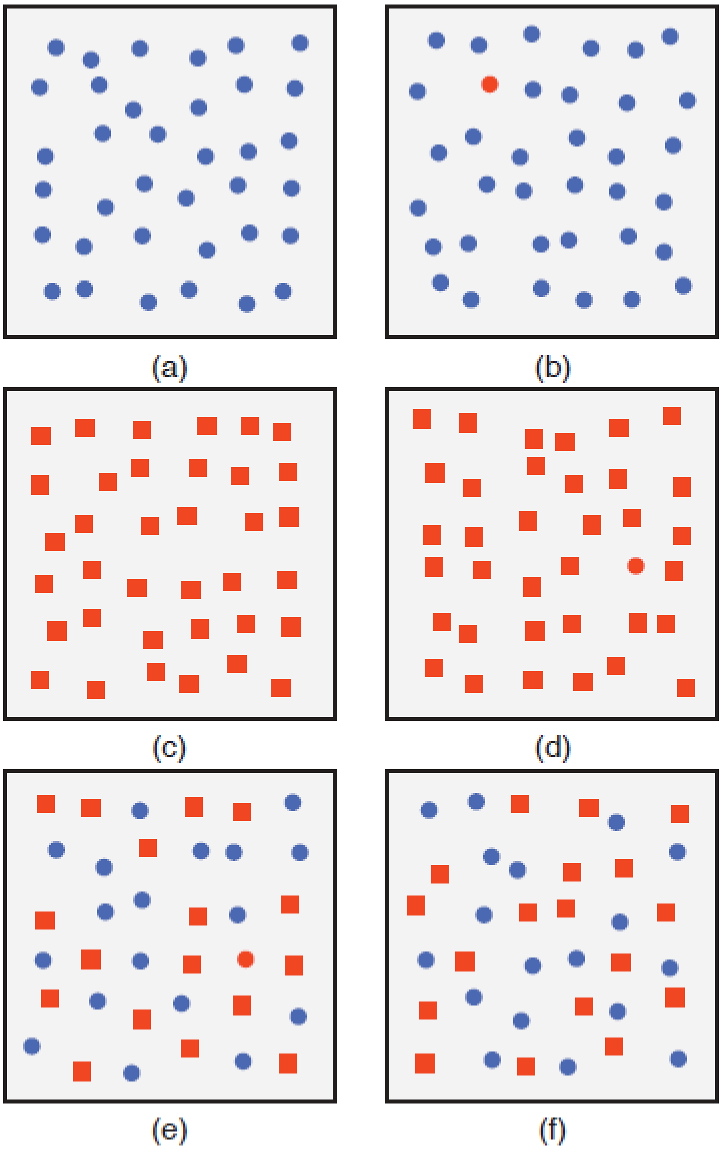

Figure 1. An example of visual search arrays from Healey and Enns (2011). Panels A and B show a colour discrimination task with the target present in B; panels C and

D show a shape discrimination task with the target present in D; and panels E and F show a combined colour and shape discrimination task with the target present in E.

Simple discriminations such as colour were originally considered to be pre-attentive as they occurred within the single first fixation (Healey & Enns, 2011). More difficult tasks, such as the shape discrimination or combined discrimination tasks are examples where Treisman's (1980) model of attention may come into play, where attention has to be allocated serially to each item within the array until the target is successfully identified.

Wolfe (1994) describes this visual search process in detail. He states that the first component of the task is disregarding certain input. In the examples in Figure 1, all items of a particular colour or shape may be eliminated. Due to our foveated vision, which limits search area to small parts of the visual field at any one time, relevant information must then be processed selectively. These limited capacity processes occur serially from one portion of the visual field to the next in order to cover the entire area (as proposed by Treisman, 1980). This is known as guided search, whereby the large scale parallel processes which are deployed to eliminate input and select more relevant areas within the visual field (Neisser, 1967) direct the limited capacity processes to the areas which are of greatest interest.

Monitoring multiple stimuli

However, with this limited capacity to process only a small portion of the visual field at a time, it is important to consider how more than one stimulus can be monitored at any given time. There has been a considerable volume of research investigating the limits of our ability to monitor multiple stimuli simultaneously. For example, Pashler (1988) produced one of the first studies to investigate change detection, where observers were asked to view a series of 10 alphanumeric characters, which after a brief pause would be presented again either in an identical format or with one character changed. Pashler (1988) found that there was no improvement to observer accuracy when they were given longer initial viewing durations until the pause between the first and second viewings was reduced to 34ms when accuracy did significantly improve. However, this effect disappeared with pauses of any greater duration. These results were cited as evidence to support a capacity-limited visual memory; Pashler (1988) reasoned that if people did have unlimited capacity, accuracy would improve at greater pause durations. In later change detection studies, similar capacity limits were discovered. Rensink (2000) performed simple visual search tasks in which observers had to detect a target object by its change in orientation; there would be one rectangle within an array of rectangles that was the correct target. Rensink (2000) reported the viewing duration required to detect change increased linearly with the number of items in the array. Observers' capacity was found to be approximately five items in an array, which supports the limited capacity proposed by Pashler (1988). It has been suggested we have a pre-attentive stage which is responsible for selecting the number of items within the visual field to be the focus of attention (Pylyshyn, Burkell, Fisher, Sears, Schmidt & Trick, 1994).

It is important then if we have limited capacity for processing to understand by what means we use gaze to attend our chosen selection of stimuli. James (1890/1950) proposed visual attention operates like a 'spotlight', and just as a spotlight selects one person on a stage, the visual spotlight selects one specific region within the visual field on which to focus, ignoring everything else. Anything that falls within the beam of the spotlight is processed more efficiently (Castiello & Umiltà, 1992). It has been suggested this spotlight may act like a zoom lens in order to encompass more items within the visual field; as the number of items increases, the lens zooms out so that more of these items fall within the spotlight. Conversely if only one item is to be attended the lens zooms in and narrows the focus of the spotlight (Eriksen & Yeh, 1985; Eriksen & St. James, 1986). This model of attention was supported strongly by Theeuwes (1991) who performed simple visual search tasks with different stimulus onset asynchronies (SOAs) which featured either rapid onset or rapid offset of a line segment near one of the letters in the array. To prevent the focusing of attention on any given location in the array, an arrowhead was centrally presented after the appearance of the display. When observers' attention was unfocused, both onsets and offsets resulted in orienting of attention to that location, regardless of where it occurred in the array. However, when the arrow was presented prior to display – thus allowing the focusing of attention at a specific location – onsets and offsets only drew attention if they occurred within the attended area. Any changes out with this area did not cause any interference to performance. Theeuwes (1991) argued that his results provided strong evidence for the idea of our attentional spotlight operating like a zoom lens, because once focused, or 'zoomed in', changes outside of the attentional sphere had no effect on observers' performance.

Alternatively, when multiple objects within the visual field must be attended, we may use a spotlight of fixed size which rapidly alternates between the objects in question several times a second (Shulman, Remington, & McLean, 1979; Tsal, 1983). The latter model has been supported by more recent research by VanRullen, Carlson, and Cavanagh (2007), who used psychometric functions to model the parallel processing strategy and the sampling strategy to human performance in simple visual search tasks. They found the prediction made by the sampling strategy (a result of rapid alternation between targets) was the best fit for the human performance data.

Regardless of the method by which a number of stimuli are monitored simultaneously, the traditional standpoint has been that humans possess a unitary spotlight that cannot be split (Posner, 1980; Posner, Synder & Davidson, 1980). It may broaden or shift rapidly, but there is only one. However, there is a branch of attention literature that questions this assumption. Early research by Pylyshyn and Storm (1988) demonstrated that at least one cognitive process (e.g. tracking a moving object) could occur simultaneously across several loci in the visual field. They reported that participants could track up to five objects in a field of 10 to detect a change in appearance, and that even when the array was displayed over a larger area than would be possible to process with a single spotlight, participants' detection ability was still very high, with 87% accuracy. Processing multiple locations in parallel was later supported with fMRI studies by McMains and Somers (2004), who found activation in the cortex which demonstrated two spotlights of attention within a single cortical hemisphere. The authors argue this is direct evidence that we can at least process low level stimuli in parallel across different loci within the visual field. In a later experiment, McMains and Somers (2005) compared the zoom lens model to the multiple spotlight model by again using fMRI to monitor processing efficiency via reaction times, event-related potential magnitudes (ERPs) and blood oxygenation level-dependent (BOLD) signal amplitude. The zoom lens model would predict deterioration in performance as the number of stimuli to be monitored increases, but this is not what McMains and Somers (2005) discovered. There were no decreases in performance, either through behavioural or fMRI measures, which the authors argue provides more strong evidence to support a multiple spotlight model of attention. This emerging model of attention is flexible and efficient, allowing observers to meet task demands by selecting relevant information regardless of their configuration in the visual field (Awh & Pashler, 2000; Müller, Malinowski, Gruber & Hillyard, 2003; McMains & Somers, 2005).

Tong (2004) explains these findings further with an example of how this splitting of attention might work in a practical setting. Citing the alleged story that Elvis Presley enjoyed watching three televisions at once, he uses this example to explain how we might split attention to perform such a task (Figure 2).

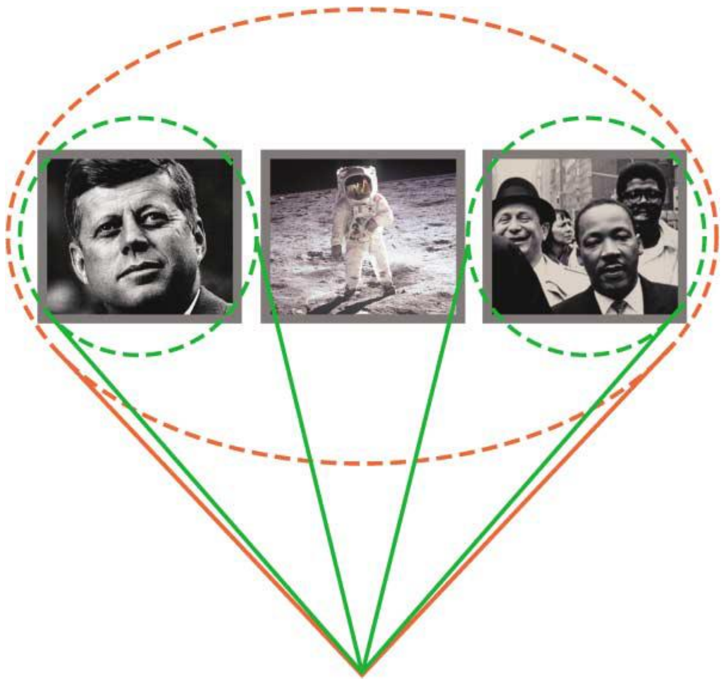

Figure 2. From Tong (2004), depicting how we may split attention across multiple screens, using the set-up of Elvis Presley's three televisions to explain McMains and Somers' (2004) fMRI findings, with the traditional unitary spotlight in orange, and the proposed split spotlights in green.

In Figure 2 shown above, the orange lines denote how a unitary spotlight attention would have to extend to accommodate the three screens. However, McMains and Somers' (2004) fMRI evidence shows that observers can attend the left and right screens without attending the centre screen (shown in green). In his summary, Tong (2004) states that McMains and Somers' (2004, 2005) use of neuroimaging to test cognitive theory resolves a debate which has been on-going for decades, and provides resounding evidence that humans can indeed attend multiple locations in space simultaneously, with little cost to performance.

Using social gaze to study attention

In previous research, visual search paradigms that do not feature real world components have been used as a means of investigating visual attention as a whole, not just the search process. One of the most renowned of these is Posner's (1980) paradigm. In this paradigm, the participant is presented with a central fixation point that has a square on either side of it. A target item appears in one of the boxes and the participant must make a key press to indicate its detection. Posner (1980) used different types of cue to determine how participants' responses were affected. Providing a congruent (correct) cue to the target's location resulted in much faster response times, but an incongruent (or incorrect) cue would result in a deterioration of performance. This paradigm became the foundation for the majority of subsequent research in to reflexive orienting of gaze as a facet of attention. Using eye trackers and modifications of the paradigm, different aspects of visual attention have been discovered, including – but not limited to – that the abrupt visual onset of an object automatically captures attention (Yantis & Jonides, 1984; Yantis &

Jonides, 1990) as does the initiation of motion by an object, whereas an already-moving object amongst static ones does not (Abrams & Christ, 2003). However, adaptations to Posner's (1980) paradigm are not limited to exploring how we respond to simple visual stimuli. A large volume of research has incorporated the basic components of Posner's (1980) original work with more complex stimuli – namely faces – to provide an avenue into studying how we allocate and orient social attention.

Friesen and Kingstone (1998) were among the first to combine spatial orienting and social cognition within one experimental framework. Rather than a centrally presented arrow cue, as in Posner's (1980) paradigm, Friesen and Kingstone (1998) used a centrally presented schematic line drawing of a face. The face would look left, right or straight ahead, before a target letter – either F or T – appeared either on the left or right of the face. Participants were told that this gaze cue was uninformative and that it would not help them with the task. However, participants still fixated on the face and were faster to respond to the target when the cue from the schematic face was congruent with target location. This effect was found across three different types of task: detection, where participants simply had to press a button when they saw the target; localisation, where they had to indicate whether the target had appeared on the left or right; and identification, where participants had to identify which letter the target had been. Friesen and Kingstone (1998) suggested that this robust effect was the result of reflexive covert attention shifts. In other words, participants were responding automatically to the gaze cue, even when they were aware doing so would not improve their performance in the task.

Driver et al.'s (1999) study followed a similar vein to Friesen and Kingstone (1998) with similar results. The eye gaze direction was spatially uninformative, as the target was equally likely to appear on either side of the face. Rather than a button press to indicate detection of the target, participants had to identify whether the target was a letter T or L. The gaze cue had no benefit for performance of the task as it gave no indication which letter would appear, or on what side of the screen it would appear. However, even though participants were reminded the gaze cues were uninformative the cues did seem to modulate participants' performance. Letter discriminations were faster when the target appeared on the side that the face gazed towards, suggesting that the participants' gaze perception resulted in spatial attention orienting to the corresponding direction. Driver et al. (1999) suggest that because participants were reminded so repeatedly these cues were uninformative, yet still derived some benefit from them, this attention shift would have to be more reflexive than intentional.

The idea of reflexive attention orienting in response to gaze cues was given further support by Langton and Bruce (1999). They conducted four experiments all using a similar letter detection task using a digitized head stimulus to provide gaze cues. Participants' success was measured with different levels of cue reliability (50% then 25% reliable); allowing extra time viewing the cue to see if participants would detect an increase in reliability to 75%; and with inverted head cues. Once again, it was found that uninformative or to-be-ignored cues produced faster detection times when they cued the target location, however this effect only held when the cues appeared 100ms before the target. Increasing the reliability did not amplify the effect in any way, but inverting the head cues disrupted it. It was also found that cueing only occurred with horizontal cues; when cues were vertical no effect on attention was found. Langton and Bruce (1999) argued that these findings were what one would expect if responding to social attention cues was a reflexive, stimulus-driven process. Furthermore, the lack of cueing afforded by vertical cues clearly demonstrates reflexive orienting is not due to participants directing their attention to certain facial features.

These studies led to a follow up by Ricciardelli, Bricolo, Aglioti and Chelazzi (2002), where they not only used gaze cues, but also non-biological directional cues in the form of arrows. They wanted to identify whether our responses to gaze cues were indeed reflexive, and what sort of cues evoked this response. The methodology was similar to the studies discussed above, using centrally presented cues with a target detection task in which they had to make a saccade towards the target. However, Ricciardelli et al.'s (2002) study had an additional step. First, there was the central fixation, then this cue turned either blue or orange indicating the direction in which the target would appear (left and right respectively). This colour cue would disappear, followed by a distractor cue, which was a either a face presented in grey-scale that would look left, right or straight ahead, or arrows that would point left or right. Participants' eye movements showed that they were less accurate when the instructional cue and distractor cue contradicted each other, with a large proportion of saccades following the distractor cue rather than the instructional cue. Both static and dynamic face cues produced the effect; dynamic cues being where the face was present when the instructional cue appeared, then showing a gaze shift. However, directional arrows did not reduce accuracy. Ricciardelli et al. (2002) suggest this shows a strong disposition in humans to imitate the gaze of others. This is one of the key behaviours in social attention, known as joint attention, where we share the attentional state of others.

Further evidence is offered in a study by Nummenmaa and Hietanen (2006), who also used gaze distractors to determine the effect on saccades. Participants were tasked with making eye movements to a target cross, which Nummenmaa and Hietanen (2006) labelled as the 'Imperative Signal'. A distractor cue was presented either 100ms prior to, or simultaneously with, the appearance of the imperative signal. This distractor cue was either a peripheral distractor in the form of a black box, or a distracting horizontal gaze cue. An example of the procedure is shown in Figure 3.

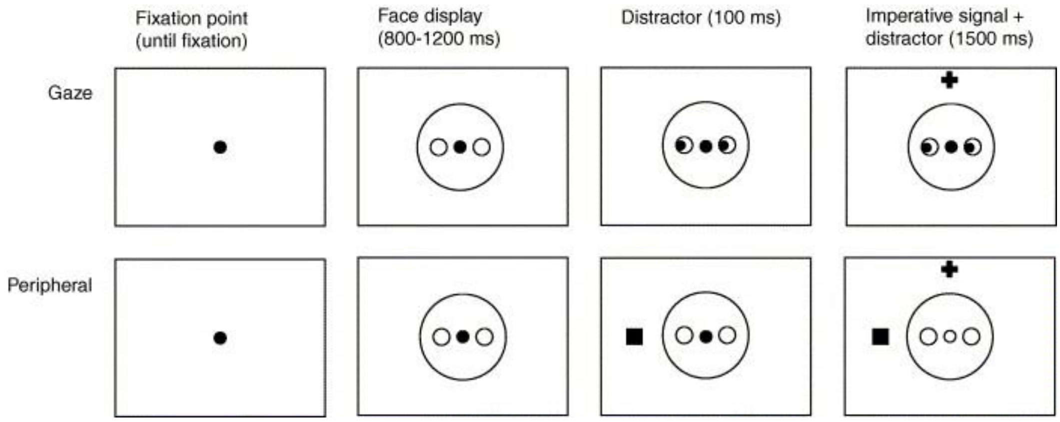

Figure 3. Trial examples from Nummenmaa and Hietanen's (2006) study showing gaze and peripheral distractors.

Nummenmaa and Hietanen (2006) found that gaze distractors caused the participants' saccades to the target to curve away from the distractor cue direction, regardless of whether the distractor was presented before the target or simultaneously alongside it. The authors argue that this provides clear evidence of an automatic activation of the oculomotor system triggering a saccade as soon as an averted gaze has been perceived. In addition, Nummenmaa and Hietanen (2006) discovered that covert orienting of spatial attention caused by distractors in the periphery influenced the curvature of saccades. They proposed this is due to the inhibition of response, where returning to a previously-searched location is inhibited for a short period of time once search has concluded by decreased activity in the neurons responsible for processing the distractor (Godijn & Theeuwes, 2002). This would suggest Nummenmaa and Hietanen's (2006) results provide strong support for bottom-up models of gaze response following gaze cueing.

Studying social attention in the real world

The above studies provide a strong theoretical basis for gaze following being a reflexive behaviour, with our attention being allocated preferentially to face or eye stimuli even when it is understood that they are unhelpful. However, in real world situations gaze cues rarely come alone. In normal social interaction gaze information is accompanied by other cues such as head direction, body posture and emotion expression. Hori et al. (2005) used a gaze cuing paradigm with photographs of real faces showing a variety of emotions. They found that when positive emotions were expressed reaction times in responding to the gaze cue given by the face were significantly faster than when a neutral or angry expression was shown. Schrammel, Pannasch, Graupner, Mojzisch and Velichkovsky (2009) also demonstrated the effects of emotion, as well as gender and gaze interaction, on attention allocation in simulated social interactions. By showing participants animated characters varying in each of these factors, Schrammel et al. (2009) demonstrated that participants' rapid facial responses, as well as their fixation times, were significantly different between angry and happy emotion presentations with participants fixating longer on angry or neutral faces. The authors suggest this is evidence that we allocate more attention to stimuli representing potential threats during social interaction. If this is indeed the case, participants could only decide if a person were a potential threat from the additional information given by the emotion presentation; something missed in traditional gaze cue studies. Although these studies are still using a laboratory-based setting, they address the fact that there is much more to any social interaction than simply a gaze cue. The studies using schematic faces (e.g. Friesen & Kingstone, 1998; Ricciardelli et al., 2002) can highlight phenomena in how our brains respond to these features, but we must always remember to interpret them with caution. It has been proposed by Sagiv and Bentin (2001) that schematic faces evoke different neural responses than normal faces. Therefore, it would be unwise to generalise results from these paradigms to live human interaction without investigation in to the stages in between.

Away from the laboratory, we use gaze cues in more complex ways. As discussed previously, we receive gaze cues in the context of a whole person, but when out in the real world these cues are part of a much richer social context than looking at a picture can portray. If we imagine ourselves walking through a shopping centre or down a busy street, we can begin to imagine where this uncontrollable, reflexive gaze following may cause problems. In these scenarios, multiple gaze cues to different locations in space are received simultaneously. Following all of these cues is impossible. Therefore, it is important to reflect this more realistic context in research. Studies based in the real world have begun to do this, and their results highlight several ways in which gaze following behaviour in the laboratory and the real world seem to differ. A good example of this comes from a real world study by Gallup, Chong and Couzin (2012a). They placed an ‘attractive object’, which was in reality a hidden camera surrounded by mirrored Plexiglas, in the middle of a busy corridor. For four days they filmed pedestrians passing the object and recorded who looked at the object and when. What they found contradicts what laboratory-based gaze cueing research would predict. When a pedestrian was faced with an oncoming passer-by looking at the object – and therefore providing a gaze cue towards it – the pedestrian would avoid looking at the object. In fact, pedestrians would only look at the object if the person in front of them looked at, which means they are following cues from head direction rather than gaze. Gallup et al. (2012a) suggest this is because there is a social cost to jointly attending an object with a stranger. This type of mutual gaze may result in potential communication, which seems to be something we attempt to avoid wherever possible. There are several other real world studies which demonstrate this apparent social cost to following gaze of a stranger, and again show it is something we tend to avoid (e.g. Gallup et al., 2012b; Laidlaw, Foulsham, Kuhn & Kingstone, 2011; Macdonald & Tatler, 2013).

Factors like the cost of potential social interaction are hard to replicate in laboratory based studies. However, it is possible to go some way to begin to replicate the more complex social scenes which are present in the real world. The research detailed within this thesis goes some way to provide a bridge between a controlled setting in which we can be confident about our manipulations, and a more realistic scenario which mimics more closely what we encounter in the real world. If the findings of the real world and laboratory-based studies discussed in the previous sections are compared, it is clear that there are considerable differences in the evidenced gaze following behaviours. There are arguments for each type of method. Kingstone, Smilek, Ristic, Friesen and Eastwood (2003) presented a comprehensive case for the use of real world paradigms when studying social attention, citing similar points to those made here, that attention studied in the laboratory may differ significantly from those demonstrated in a real situation. They warn that by accepting lab-based research as demonstrative of real-world behaviours, the study of attention may be warped by “fundamental misunderstandings of the principles of human attention and behaviour” (p.179; Kingstone et al., 2003). This is a powerful and important point. The authors cite traditional visual search paradigms that tend to use simplified stimuli, such as geometric shapes, the results of which have led to several proposed models for visual search. This is discussed further by Tatler, Hayhoe, Land and Ballard (2011), who are also sceptical of how much it is possible to infer from these types of paradigms. They discuss the traditional models of gaze allocation, which tend to be based on data from static picture viewing, the most prominent of these models being image salience. Again, this is a model based on behaviour that is not always seen in real world gaze, and which cannot explain the majority of behaviour in natural tasks (e.g. tea making; Land, Mennie & Rusted, 1999; or making a sandwich; Hayhoe, Shrivastava, Mruczek & Pelz, 2003). These models do not necessarily fit results from real world visual search data that suggest search is affected not only by primitive features such as shape, but also by complex attributes like social context or significance (Kingstone et al., 2003).

With these concerns in mind, it is important to consider research that goes some way to bridge the gap between the controlled setting of the laboratory and the richer context of the real world. There is substantial evidence to suggest that when observers look at a scene, for example a photograph, they preferentially fixate on people within the scene over anything else. This was first documented by Yarbus (1967) who used a painting called ‘An Unexpected Visitor’ by I. P. Repin to study how participants moved their eyes when viewing the painting (Figure 4).

Figure 4. Left: the original painting, Middle: eye movements when asked to freely view the scene; Right: eye movements when asked to rate the material wealth of the family.

Yarbus (1967) asked the same individual to view the scene several times with different tasks given for each viewing. As can be seen in Figure 4, the scan path across the scene varies considerably when the viewer is asked to freely view the scene from when they are asked to assess the material wealth of the family. Other tasks included estimating the ages of each person in the scene, deciding what the family had been doing before the visitor arrived, and memorising either clothes worn or position of people and objects (Tatler, Wade, Kwan, Findlay & Velichkovsky, 2010). Yarbus' (1967) work highlighted the importance we place on viewing faces to understand socially complex scenes. He hypothesised that the decision to attend gaze information is influenced by the task being performed; that fixations on the face, and gaze-following, would be modulated by the social content and complexity of the scene presented and the activity within that scene (Birmingham, Bischof & Kingstone, 2008). Given the previously discussed anatomy of the eye, which makes it ideal for non-verbal communication, and the fact that humans have evolved an internal reward mechanism that reinforces the preference for social stimuli (Leder, Tinio, Fuchs & Bohrn, 2010), it seems intuitive that faces and eyes will be preferred when trying to disambiguate a scene.

People-Preference

Research has continued to support the preference for people over all other types of stimuli. Fletcher-Watson, Findlay, Leekam and Benson (2008) carried out a study with a preferential looking paradigm in which two scenes were presented side by side. These scenes were balanced for complexity, with the only difference being that one scene had a person in it, and the other did not. In one condition, participants were asked to determine the gender of the person in the people-present picture. In the other, participants were given no instruction to monitor free-viewing eye movements across both person-present and person-absent scenes. Participants showed strong bias to study the person-present scene rather than the person-absent scene, and this effect persisted in the free viewing condition. This bias was measured by both first fixation and total number of fixations, and the person-present scene received a much higher proportion of both. There was no difference in the time spent looking at the background of either scene, which demonstrates the increased fixation duration on the person-present scene is solely due to the time spent looking at the person in that scene. Fletcher-Watson et al. (2008) argue that the strong preference for the person-present scenes affects the landing position of the first saccade, even within the first 100ms after scene presentation. This rapid attentional capture by high-level stimuli is an effect demonstrated in several other studies (Herschler & Horstein, 2005; Kirchner & Thorpe, 2006; Smilek, Dixon & Merikle, 2006), and in the case of Fletcher-Watson et al.'s (2008) study, that high-level stimulus would be the person in the person-present scene. Their study was the first to demonstrate rapid attentional capture within a spontaneous looking condition, which they interpret as being due to our innate need to very quickly process complex, socially-relevant information.

An additional feature of Fletcher-Watson et al.'s (2008) study worth of note is that face stimuli were presented within a whole person in a real word scene – participants were shown photographs rather than digitally created images. While this may seem intuitive, it is an important distinction as other studies using paradigms where only a face was presented have failed to elicit first saccades to faces within a search task (e.g. Brown, Huey & Findlay, 1997). Even in the free viewing paradigm, participants landed 15% of first fixations on the face of the person, and these fixations had no increased cost to preparation time in comparison to fixations on other areas. It is possible that a face is more readily recognised when supported by other contextual clues such as the presence of a body, and Fletcher-Watson et al. (2008) found the majority of first fixations were on the body rather than the face. They posit that while the face is a powerful stimulus, its presence is small within the viewing angle in comparison to the body. This would explain why most first fixations are commonly directed to the body with a very high likelihood of moving to the face on the second fixation. This pattern of fixations was evident in both the free viewing and gender-discrimination conditions. While Fletcher-Watson et al.'s (2008) results demonstrate a highly focused distribution of fixations on the face of a person in the scene, this only occurs after a first fixation on the body. As a result, the authors argue that although the pattern of fixations suggest the aim of moving gaze towards the highly socially-relevant stimulus of the face, this movement occurs in stages and therefore the whole figure – face and body – are involved in the initial ultra-rapid processing of visual information.

Even when people are made relatively inconspicuous within a complex scene, observers' gaze will still be drawn to them. Zwickel and Vō (2010) asked participants to look at a number of full colour scenes with instructions to view them as if they were looking through photographs. These complex scenes (shown in Figure 5) featured a person, but they were not made salient either by instruction or low-level features. In order to produce natural eye movements, object salience was controlled for by keeping the objects constant within scenes, changing only the orientation of the person.

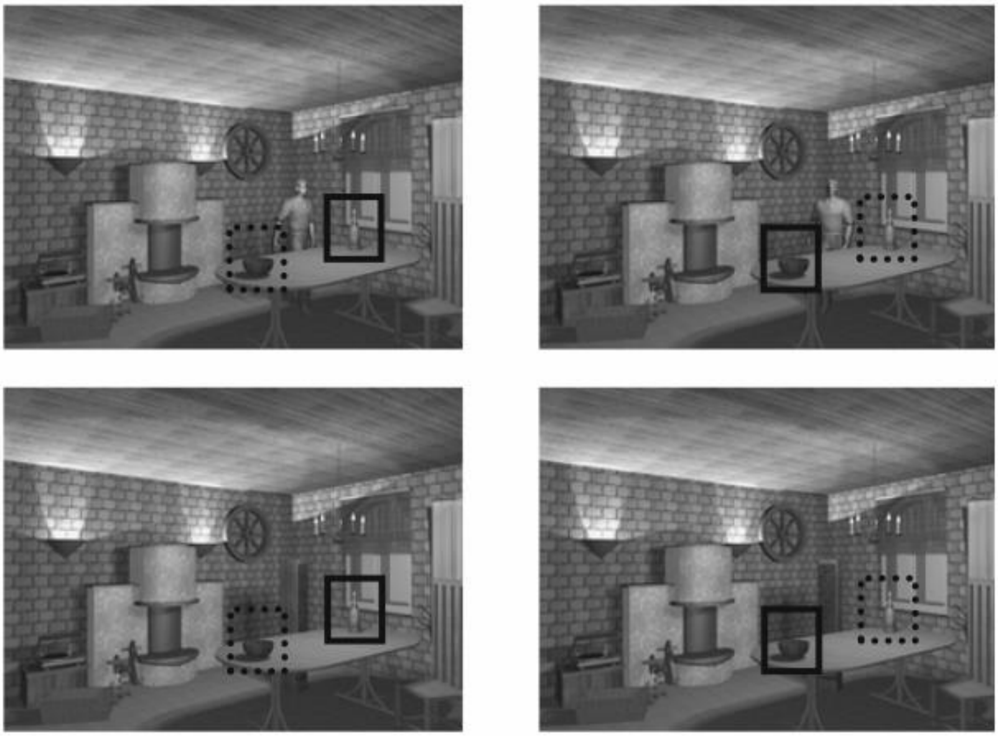

Figure 5. Example stimuli from Zwickel & Vō (2010). In each image, the object cued by the person or inanimate object is highlighted by a solid black box, and the uncued object is highlighted by a dotted black line.

Zwickel and Vō (2010) also wanted to ensure gaze cueing was not the result of clear orientation and so replaced the person with a loudspeaker for half of the scenes, which was chosen due to its similar size as the person within the viewing angle. Furthermore, both provided directional information by their orientation. In person-present scenes, the head was too small to accurately determine eye gaze direction, and so other directional cues from the head and body can be used to guide gaze.

The main aim of Zwickel and Võ's (2010) study was to determine whether a person presented in a complex scene would be preferentially fixated even when they were not central to the task, and they found that this was indeed the case. Participants preferentially fixated the head region of the person in the scene over other stimuli, and as a result their gaze was directed faster, more frequently and for a longer duration to the object cued by the person's gaze. This gaze-following was only seen in the person-present scene, and not when the person was replaced with a loudspeaker, which the authors argue is evidence that repeatedly shown inanimate objects do not lead to gaze cueing. As stated, participants preferentially fixated the person's head, even though this was very small in comparison to the body. Again, the distinction is important – fixating on the body did not result in following the direction cued by body position. Zwickel and Võ (2010) determine that fixations on the person are not simply a result of extracting orientation information, as if this were the case there would be no need to fixate the head. The clear preference for using head direction as a cue suggests that we are driven to ascertain what social information is conveyed by head direction when viewing complex scenes.

Attention is drawn to the eyes

It has been established that people are the preferred stimuli within scenes, and that within a person, gaze is drawn to the head as it contains more socially relevant stimuli. There is a host of research that has looked at narrowing this preference further by studying which features of the head are most highly preferred when discerning features is possible. As has been previously described, the physiology of the eyes makes them ideal for functioning as communicative tools. Emery (2000), and Langton, Watt and Bruce (2000), have described eye gaze as having special status compared to other directional cues (e.g. head and body position, or pointing). Birmingham, Bischof and Kingstone (2009) studied the selection of stimuli as a means of examining social attention. As in studies mentioned previously, Birmingham et al. (2009) used realistic stimuli in order to elicit natural fixations. Their aim was to determine whether eyes were preferentially fixated over other directional stimuli; in this case, arrows. Participants were asked to freely view a number of scenes in which both a person and an arrow were present. When both types of directional stimuli were present in a scene the majority of fixations were on the eyes of the person with very few fixations on the arrow. First fixations were equally likely to fall on the eyes, heads or text within a scene, but almost never on the arrow, which would suggest that arrows are seen as being relatively low on the list of informative stimuli. When viewing time was increased, participants continued to show preference for the person – particularly their eyes – and continued to ignore arrows. Even when the arrow was much larger than the people in the scene (Figure 6), they were never fixated first and had a very low proportion of the total number of fixations.

Figure 6. Stimulus from Birmingham et al. (2009) showing the original scene on the left, with fixation data added on the right. This example clearly shows an arrow much larger than the people present in the scene.

Birmingham et al. (2009) argue that their results are consistent with evidence that the human brain has specific mechanisms preferentially biased to processing eyes (as shown by Pelphrey, Morris and McCarthy (2004) amongst others). This would suggest that while arrows may convey a directional cue as accurately – if not more so – than a gaze cue, they are not allocated the same level of priority by the attention system when viewing complex or realistic scenes. Although arrows do elicit reflexive attention shifting in Posner-type cueing paradigms, Birmingham et al. (2009) would suggest that in the real world, arrows are only attended in circumstances where they are task relevant, for example when taking an exit on the motorway. Arrows are not the only stimuli to receive a lower priority ranking. The results would suggest that people have an inherent hierarchy of preferential social stimuli – eyes are most preferred, but when this information is unavailable head direction will be used, and if this information is unclear fixations will be made on the body. This would support the model of social attention proposed by Perrett, Hietanen, Oram, Benson and Rolls (1992), which puts gaze at the top of the social attention cue hierarchy, followed by head direction and then body position.

It is clear then that eyes are the preferred stimuli over all others within the context of a person. The question then is to explain why eyes and gaze are so important to our understanding of the world around us. From examining the development of gaze-related behaviours in infancy, it is clear that in infancy gaze is important for learning about the world. Similarly, adults use gaze information to disambiguate complex real world scenes, and preferentially fixate eyes even when given no task concerning people within a scene. Birmingham, Bischof and Kingstone (2007) proposed that together, this evidence suggests an inbuilt understanding of eyes being highly informative. They tested this hypothesis by giving observers both social and non-social scenes. Half of their participants were told they would be asked to recognise scenes in a test session (the Told group), whereas the other half were told there would be a memory test after they had freely viewed the scenes, but the content of the memory test was not disclosed (the Not Told group). Following their viewing of the scenes, participants were shown a series of scenes, some they had seen before and some which were completely new, and were asked to identify which of the scenes they had seen before. The results showed a clear preference to fixate eyes in people present scenes when observers were encoding them to memory than when they were asked to freely view the scenes. What is particularly noteworthy in Birmingham et al.'s (2007) results is that the Not Told group fixated on eyes much more strongly during the testing session than during the viewing session, which the authors suggest is evidence that eyes are seen as the most informative stimulus within people present scenes.

The Research Aims of this Thesis

Together, all this information establishes a very strong theoretical foundation for investigating social gaze cues. From this research, it is understood that gaze is one of the most useful tools for investigating attention, and it seems likely that when multiple stimuli must be tracked at once a split spotlight of attention is used. It is clear that there is a strong preference to look at people within scenes, which is most likely because they are seen as highly informative stimuli. Within a person, eyes are at the top of the informative-stimuli hierarchy, followed by face cues or head direction and then body direction. Humans have evolved to recognise eyes as highly informative cues both for non-verbal communication and to aid in verbal communication, and our eyes' physiological structure makes them ideal for this role. The process of following gaze cues is so ingrained some studies suggest it cannot be inhibited, even when it is known that these cues will not be helpful to the task being performed. However, differences between gaze behaviour in laboratory studies and real world studies are beginning to emerge, raising questions about the ecological validity of laboratory studies using simplified geometric visual search arrays, or schematic or digitized faces. Nevertheless, to understand gaze in the real world, some means of studying it in a controlled setting must first be established.

Research Aim

The research contained in this thesis aims to build on the foundation of social gaze research, and to provide a bridge that goes some way between traditional laboratory studies and real world paradigms. Here, a new visual search paradigm has been created that uses photographs of real people in real scenes, depicting natural and uninstructed gaze cues towards objects. This paradigm provides a template that can be used to explore other facets of how social gaze is processed and utilised. The following provides an outline of how this will be tested over the course of the chapters contained in this thesis.

Thesis Outline: Single Gaze Cues

Chapter Two tests this paradigm for the first time presenting observers with a single gaze cue. This cue is either congruent or incongruent with the target object location. Chapter Two is designed to provide a more direct comparison between this new methodology and previous Posner-type tasks. Chapter Three introduces different types of task instruction to the experiment by manipulating how useful the cues presented are perceived to be by participants. This chapter presents two studies: in the first study participants are told the cues are not relevant to the current task and they are to be ignored, in the second study participants are told the cues may be useful in finding the target object. The first study fits more closely with the types of task instruction given in previous research, whereas the second study explores a relatively undocumented approach to task instruction in Posner-type paradigms. Where Chapters Two and Three compare the effects of gaze cue condition on participant performance, Chapter Four compares the effects of the different types of instruction on participant performance in each measure of search. This omnibus chapter examines each search measure in detail and discusses how the instruction given to participants regarding the purpose of person presence in the scene may affect performance within each gaze cue condition.

Thesis Outline: Multiple Gaze Cues

Providing participants with multiple spatially uninformative gaze cues that are simultaneously-presented is introduced to the task in Chapter Five. The aim of this chapter is to determine how observers respond when they are given multiple cues that may provide conflicting gaze information. This chapter follows the same general methodology of that employed in Chapter Two: participants are given no instruction regarding person presence, and are just asked to find the target object as quickly as possible. The effects of gaze cues that are incongruent and congruent with target location are explored, in addition to a people absent gaze cue condition and a conflicting gaze cue condition where each person in the scene cues a different object. Chapter Six presents two studies, each testing a different task instruction that is designed to manipulate the perception of people presence to be either helpful or unhelpful. In this chapter the effects of gaze cue condition are examined separately within each study. The second omnibus chapter is Chapter Seven, which compares the previous three studies in each performance measure. This chapter is designed to determine the effects of instructions regarding people presence on participant behaviour in the visual search task. Rather than investigating how the different types of gaze cue impact participant performance, this chapter provides a quantitative comparison of the effects of instruction within each gaze cue condition. To explore the effects of manipulating gaze cue sender reliability, Chapter Eight uses only the conflicting gaze cue condition, which is adapted so that one person is always more reliable than the other, predicting target location on 70% of trials. The effects of reliability are explored with and without the addition of instructions regarding person presence.

Finally, Chapter Nine pulls together the results of each of these studies and discusses the implications of their findings. It compares the differences in participant performance when provided with single or multiple cues, how instruction regarding people presence impacts on performance, and what the effects of manipulating the reliability of gaze cue senders are. This chapter discusses how these findings contribute to the body of social attention research and how the use of realistic stimuli changes predictions about real world gaze behaviour. This chapter also identifies how this paradigm might be deployed in future research to further our understanding of social attention both in typical and atypical populations.

Chapter Two

Updating Posner-type Paradigms: Using realistic environments to explore the effects of gaze cues on visual search

Introduction

The ability to search and locate a single item within our environment is a key component of many everyday tasks. Finding the desired groceries in a supermarket, spotting a friend in a crowd, or finding a clean mug for making coffee are all examples in which visual search is employed. While these real world examples may also utilise other factors such as knowledge of the layout of the supermarket, or memory of where recently cleaned crockery has been stored, understanding the underlying principles of how search is conducted requires breaking down these complex processes so that they can be studied in closer detail. This chapter describes a new search task that is designed to study search in a more simplified manner.

Posner's (1980) paradigm – where a centrally presented cue that is either congruent or incongruent with target location – was initially used to explore low-level aspects of visual attention. For example, Yantis and Jonides (1984, 1990) used the valid/invalid cue structure to determine whether the abrupt onset of a stimulus resulted in a reflexive attention shift. Through several experiments varying participant attentional readiness, the appearance of a cue before, during or after the test display, and predictive validity amongst other variables, it was determined that abrupt onsets did indeed trigger a reflexive shift of attention. Jonides (1981) also tested the effects of this type of cuing in the periphery of the visual field. He found the same reflexive attentional shifts to areas highlighted by the peripheral cues, even when participants were told to ignore them.

As Posner (1980) had predicted, examining orienting in a reductionist manner was quickly recognised as a valuable way to understand how we respond to gaze cues. With the direction of gaze being easily identifiable as a result of the dark iris and pupil contrasting with the white sclera, a gaze cue provides much the same directional cue as the arrows used in early Posner-type studies. Among the first to adopt this new method of exploring gaze were Friesen and Kingstone (1998) who presented observers with a schematic line drawing of a face as opposed to a central fixation cross. The eyes would either provide a neutral cue, where they looked straight ahead, or gaze to the left or right, providing either a valid or invalid cue to the target's location. As in previous Posner-type studies, Friesen and Kingstone (1998) found that valid gaze cues resulted in faster responses to the target than invalid or neutral cues. Crucially, the authors told participants that the gaze cues were not helpful and would not predict the target's location. The pervasive facilitation in performance produced by valid gaze cues was cited by the authors as evidence of reflexive orienting of attention in response to gaze cues; a conclusion that has been supported by later research (e.g. Driver et al., 1999; Ricciardelli et al., 2002).

Ricciardelli et al. (2002) had argued that the reflexive orienting response was unique to the biological stimuli of gaze, and would not occur for other stimuli. This received support from Friesen, Ristic and Kingstone (2004) who used counter-predictive arrow and gaze cues to study reflexive orienting, and found that only gaze produced reflexive responses; the orienting to areas cued by arrows was judged as volitional. Regardless, the unique status of gaze was refuted by Kuhn and Benson (2007), who found shorter saccade latencies and more saccades toward the target in valid cue trials regardless of whether the cue provided was an arrow or a schematic line drawing of a face. Other research argued that neither gaze nor arrows trigger a rapid reflexive orienting response, stating that instead these shifts in attention occurred only when cues were presented for longer durations and overlapped with the presentation of the target (Green, Gamble & Woldorff, 2013).

In Posner-type tasks, cues are always presented centrally. Participants' central fixation is already on the centre of the face, so no eye movements are required to receive gaze cues. However, it is debatable whether this presentation of the face in the exact centre is the best representation of following gaze in the real world. Tatler (2007) explored the clear tendency of observers to fixate in the centre of a scene when viewing it on a computer monitor. He cites that this is often assumed to be the result of the main content of images usually being presented centrally – for example if taking a photograph of a person we normally centre them within the frame. Alternatively, it could be assumed this tendency is the result of using central fixation points prior to the scene appearing, then making saccades with small amplitudes. However, the results from Tatler's (2007) study could not be explained by either of these assumptions. The stimuli used balanced the position of scene content, thus the assumption of the most interesting components of a scene being in the centre is immediately discredited as an explanation for central bias. There was also no evidence found to support small saccade amplitudes rather than large amplitudes. Since the central bias tendency persisted regardless of how the image was composed or the task given to participants, Tatler (2007) posed three possible explanations for bias, including potential optimal viewing position, convenience, or tendency to bring the eye back to the centre of its orbit. The salient point is that it is not clear why the tendency to fixate on the centre of a screen occurs, but it persists despite various scene manipulations. This raises the question of whether the central bias tendency could also affect how participants respond in Posner-type tasks where the directional cue is presented in the centre of the screen. Are the facilitation effects observed solely due to the cue provided, or are there also effects of this bias? Using this type of paradigm means it is impossible to tell.

Furthermore, whether these cues are from a schematic line drawing of a face, or a digital image of the head and shoulders of a person, they do not mimic how we encounter gaze cues in the real world; alongside the context of a body (Birmingham & Kingstone, 2009). It is questionable then whether conclusions drawn from studies that use stimuli that are unlike how we would receive a gaze cue in the real world can be used to extrapolate natural gaze behaviour. ‘Gaze’ is not solely a definition of where the eyes are pointing. Research shows that when interpreting gaze direction we also use visual information from head direction and even body orientation (Hietanen, 2002). Langton (2000) used a Stroop-type interference task to determine whether head direction was utilised in determining gaze direction. Participants were tasked with making key-press responses indicating in which direction the gaze cue was given. This cue would feature either congruent head and eye cues where both pointed in the same direction, or conflicting cues where the head and eyes would cue opposite locations. Participants were asked to follow either the head or the eye cue, and results showed that a conflicting head cue slowed responses to the same extent as a conflicting eye cue. Therefore, it can be concluded that gaze is not solely about where the eyes are pointing – although this is an important feature – we also use head direction to analyse where attention is being allocated. In order to properly explore responses to gaze cues then, cues must be presented that mimic natural behaviour. The cue must occur in the context of a whole body, and involve both head and eye movement, not just a shift of the iris and pupil while the head remains stationary.

The current study aims to combine the main principles of Posner-type tasks within a more realistic environment. Here, a single non-predictive cue is provided by a person sitting behind a table, giving a gaze cue that encompasses both the eyes and the head. The cue is provided in a realistic context: each scene encompasses the visible upper half of the person's body as they sit behind the table. The cue is not in the centre of the screen – to overtly seek out gaze an eye movement must be made. The target position has been carefully counterbalanced so that it never appears in the centre; instead it is equally likely to appear on the left or right side of the table. Like Posner-type tasks then, the current study uses a central cue toward a peripheral target. However, by using photographs of real scenes, the task takes place in a more realistic environment, thus providing a direct comparison by which it can be determined whether the facilitatory effects seen in traditional Posner-type studies are found in more natural environments.

To provide a baseline for later studies explored in the follow chapters of this thesis, the current study provides participants with no instruction regarding the presence or absence of a person within the scenes, instead the search task is explained in detail and they are asked to find the target object as quickly as possible. In the Posner-type tasks discussed, instructions regarding gaze cues are mixed. Some tell participants only that the cues are irrelevant to the task (e.g. Driver et al., 1999; Friesen & Kingstone, 1998; Friesen & Kingstone, 2003a), others told participants explicitly to ignore the face in the scene (e.g. Ricciardelli et al., 2002), while others still gave no instruction regarding the face (e.g. Friesen, Ristic & Kingstone, 2004). All of these studies use centrally-presented schematic line drawings or digitised images of faces. In the current study where the search environment is quite different, I felt it was important to first gauge how participants used gaze cues provided when they could freely decide to use or ignore them as this would provide a guide to how participants utilise this information when their only task is the search. Making the person salient by discussing their purpose adds another layer to the task, and this will be explored in later chapters.

Method

Participants

A total of 43 people (15 male) were recruited for participation in this study. All had normal or corrected vision and were naïve to the purposes of the study. Level one and two undergraduate students received course credits for participation; anyone not eligible for course credit was paid £2.

Materials

Experimental scenes were created using ten different sets of everyday objects. Each scene featured one of the ten sets of 15 everyday items arranged on a table top. It was found through pilot testing that an array of 15 objects was large enough to prevent the task becoming too easy, but not so many that the scene became overly-crowded, which could have made the target of the provided gaze cue less clear. Within each scene one item was designated the target and another designated the distractor. These items were always on the opposite sides of scene centre, so that looks to these items are unlikely to be due to the typical human bias to look near the scene centre irrespective of content (as discussed in Tatler, 2007). If imagining the scene split into thirds, the target and distractor would always be positioned in the left-most and right-most thirds, never in the centre. The target was equally likely to appear on the left or right side of the table. An example of the scenes presented are shown in Figure 7.

Figure 7. An arrangement of an experimental scene showing the two photographs for one arrangement of objects on the table top. In this arrangement the target is the Filofax and the distractor object is the earmuffs. Box A shows the individual looking toward the target and Box B shows the individual looking toward the distractor.

In Figure 7, Box A provides an example of the *congruent* gaze cue condition, where the person cues the target object. Box B shows an example of an *incongruent* gaze cue condition where the person cues the distractor. For each set of 15 objects, the objects would be laid out on the table with the target on the left hand side of the scene and the distractor on the right hand side (depending on the version of the experiment, either item may be highlighted as the target object). This would be photographed once with the person looking at the target (a *congruent* cue), then again looking at the distractor (an *incongruent* cue). The array would then be rearranged so that a new pair of target and distractor objects would be positioned on either side of the table. All the other objects within the array would be rearranged so that they were not in the same position as in the previous shots. The scene would then be photographed again, once with the person looking at the target and then at the distractor. This would be repeated twice more, with different target and distractor items in each arrangement. Finally, the person would move out of the scene and the objects rearranged again and photographed. A total of 10 person-absent scenes and 80 person-present scenes were created.

Full counterbalancing required a large number of experimental scenes, and creating object arrays of 15 different objects for each one would have been exceedingly difficult. For this reason, it was decided multiple arrangements would be used to allow repeated use of the same object sets, provided each arrangement used different target and distractor objects (which were randomly selected) to prevent any learning effects. Therefore, to ensure full counterbalancing, each participant saw each set of items three times – two person-present scenes and one person-absent scene – but each time with the objects arranged differently and with a different object as the target. Within each participant, the trials were counterbalanced so that the participant was never asked to look for the same object twice within the same array of objects; that the target appeared on the left of the screen in 50 trials, and on the right in 50 trials; and that there were an equal number of congruent gaze cues directed toward the left of the screen as there were congruent gaze cues directed to the right of the scene.

Two versions of the experiment were created so that counterbalancing could also be completed across participants. The second version of the experiment effectively operated as a mirror of the first: for example where version one of the experiment would show participants Box A of Figure 7 and ask them to look for the Filofax (receiving a congruent cue), version two would show participants Box B when asking them to look for the Filofax (an incongruent cue). Essentially, when version one provided participants with a congruent cue, version two would provide an incongruent cue, and vice versa.

Eye Tracking

Participants' eye movements were tracked using an SR Research EyeLink 1000 eye tracker with a sampling rate of 1000 Hz, using pupil tracking and corneal reflection. The tracker was desk-mounted, sitting below the computer monitor and used to track a participant's dominant eye. The participant's head was kept stable throughout the experiments using an adjustable chin and forehead rest. Stimuli were presented on a 19-inch CRT computer monitor with a resolution of 1024 x 768 pixels. The Experiment Builder software developed by SR Research was used to run the experiment. Calibrations performed using the EyeLink 1000 were accepted if the average spatial error was less than 0.5 degrees and the maximum error was less than 1 degree over the 9 calibration points.

Procedure

A single-point calibration check was performed before each trial began. The name of the target object for the trial was presented on a grey-scale background for 500 ms. Most target object names were high frequency words, but to control for variation in participant vocabulary, a 500 ms presentation time was used following mean naming time of written words across low to high frequencies established as 546 ms by Schilling, Rayner and Chumbley (1983). This was followed by the presentation of a blank screen for a further 500 ms. In experiments where participants freely view scenes, it is common practise to include the presentation of a white noise mask for 500 ms following scene presentation to prevent interference between trials (e.g. Tatler, 2007; Tatler & Vincent, 2008; Tatler & Vincent, 2009). In the current study a 500 ms blank screen was shown prior to scene presentation to prevent interference from any residual word processing following presentation of the target word. After this blank screen presentation, the visual search scene appeared. This trial procedure is shown in Figure 8.

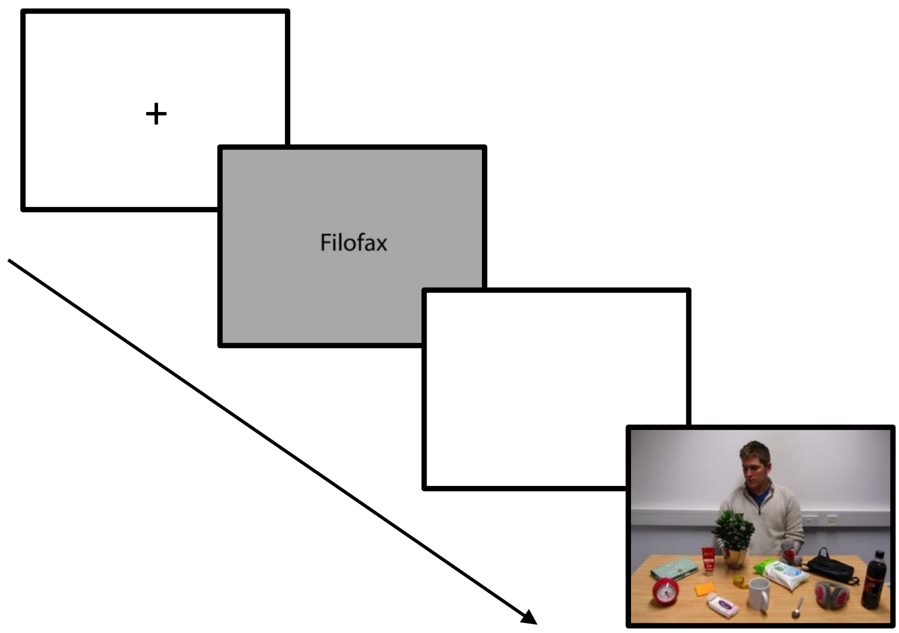

Figure 8. An example of a single trial procedure.

To indicate they had found the target, participants were asked to press either of the trigger buttons on a Microsoft Sidewinder gamepad – whichever they found most comfortable to use. Scene presentation ended with the button press or after 10 s had elapsed with no response. Each participant saw a total of 100 scenes: 20 person absent scenes and 80 person present scenes. In both versions of the experiment, all ten person absent scenes would be presented twice, with participants asked to search once for the target object and once for the distractor object. A total of 40 person present scenes were each shown twice, again with participants required to search once for the target and once for the distractor. The difference between versions one and two of the experiment was simply which of the two photos within an arrangement (as shown in Figure 7) were presented.

Participants were given no instruction regarding the presence or absence of a person in the scenes. They were given a brief description of what would happen in each trial and simply asked to find the target object as quickly as possible.

Data Analysis

Data were analysed using linear mixed effects models (LMMs) in the lme4 package (Bates, Maechler & Bolker, 2011) within the R statistical analysis environment (R Development Core Team, 2011). For logistic models, the lmer() function returns z-values and estimated p-values for each effect. For linear models, the lmer() function returns t-values without the associated p-values. In these models, we consider any effects for which the t-value is greater than two – that is effects larger than twice their standard error – as reflecting a significant effect (as in Kleigl, Hohenstein, Yan & McDonald, 2012). In all models gaze cue condition was included as a fixed effect, and simple effects were reported. Each variable was explored by an initial model with three levels in the fixed effect (person absent, incongruent gaze cue, congruent gaze cue) using the person absent condition as a baseline to which other conditions are compared in the simple effects. Follow up models were run in order to explore differences between the two gaze cue conditions (therefore excluding the person absent condition), in which the congruent condition was used as the reference condition. In any model where data were skewed, a logarithmic transformation was used to generate a normal distribution for analysis. In the results section it is noted if transformation was required.

In all models the random factors of participant and scene were included. Where possible the maximal model was used in which intercepts and slopes for the fixed effect of gaze cue condition was allowed to vary over both of the random factors (Bates et al., 2014). However, such maximal models often fail to converge without large amounts of data and so random effects structures were simplified whenever necessary. The simplification process was first to remove the parameter estimating the correlation between the slope and intercept of the fixed effect. If further simplification was required the slope for the fixed effect was removed from the scene random factor (this is an intercept-only random structure for scenes), but retained across participants. The final simplification step was to run a model with intercept-only structure for both random factors. In the following analyses the model with the most complicated random effects structure that converged was reported.

In the analyses that follow variables are considered that reflect two different stages of search (see Malcolm & Henderson, 2009; Spotorno, Malcolm & Tatler, 2014): search initiation (first saccade latency, first saccade direction, first saccade end point accuracy), and scene scanning (time to first fixate target, scan path ratio). Overall search behaviour is analysed in terms of the response times of participants to press the button, terminating search. Behaviour with respect to the individual pictured in the scenes (number of looks at the person) is also considered.

Results

Search Initiation

Search initiation was explored using three separate measures. First the time taken to launch the first saccade after the appearance of the scene (first saccade latency) was evaluated. The direction of that first saccade was considered by measuring the proportion of trials in which the saccade was directed towards the target object. A saccade launched in the direction of the target was defined as any saccade for which the angular direction of the saccade was within 22.5 degrees of the angular direction toward the centre of a bounding box placed around the target object (as in Spotorno et al., 2014). Finally, the distance from the landing point of the first saccade to the centre of the target (end point accuracy) was assessed. Together, these three measures illustrate the time spent processing scene information before initiating overt search and the subsequent accuracy of the initial eye movement of that search, both in terms of its direction and how close it brought gaze to the target object. Each of these measures considered data from all trials, irrespective of whether the target was fixated later in the search or not.

First saccade latency analysis used data from all trials, and these data required two transformations. The data presented a small number of very short latencies, which most likely were the results of pre-emptive eye movements beginning before the appearance of the scene. Very short latencies were defined as being less than 100 ms. A total of 539 trials featured a very short latency (12.5% of the total number of trials), and these very short latencies were removed. The remaining data underwent logarithmic transformation to generate a normal distribution. The time taken to launch the first saccade after the scene appeared (first saccade latency) was significantly reduced when there was a person in the scene, regardless of whether they were giving an incongruent, $\beta = -0.011$, $SE = 0.004$, $t = -2.4$, or congruent gaze cue, $\beta = -0.016$, $SE = 0.004$, $t = -3.3$ (Figure 9). The follow up LMM found there was some further decrease in first saccade latency in the congruent gaze cue condition as compared to the incongruent condition, $\beta = 0.004$, $SE = 0.003$, $t = 1.2$, but this did not reach significance.

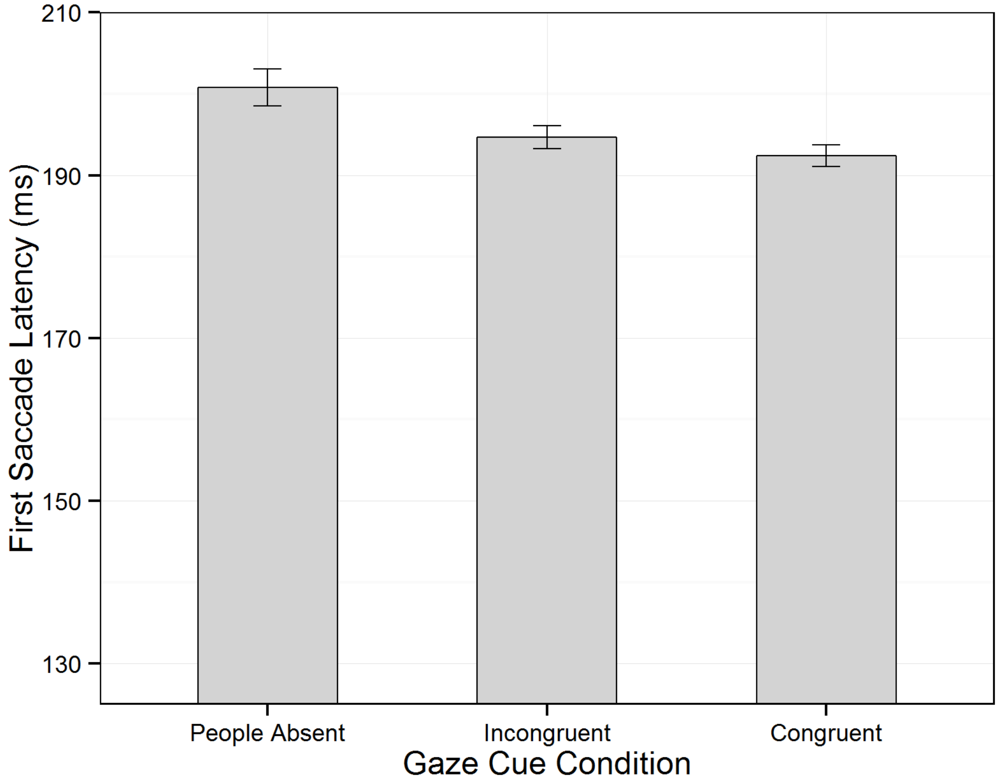

Figure 9. First saccade latency in each of the three gaze cue conditions. Error bars show standard error across all data samples.

The direction of the first saccade showed much greater variation across the three gaze cue conditions. Again, data from all trials were used in this analysis. Both person present conditions produced significantly more first saccades directed towards the target after the scene appeared than in the person absent condition (incongruent: $\beta = 0.112$, $SE = 0.020$, $t = 5.516$; congruent: $\beta = 0.124$, $SE = 0.020$, $t = 6.155$). The proportion of first saccades directed towards the target across the three gaze cue conditions is shown in Figure 10. The follow up LMM compared the proportion of first saccades between person present conditions. However, as in the first saccade latency measure, there were no significant differences between the person present gaze cue conditions ($t < 1$).

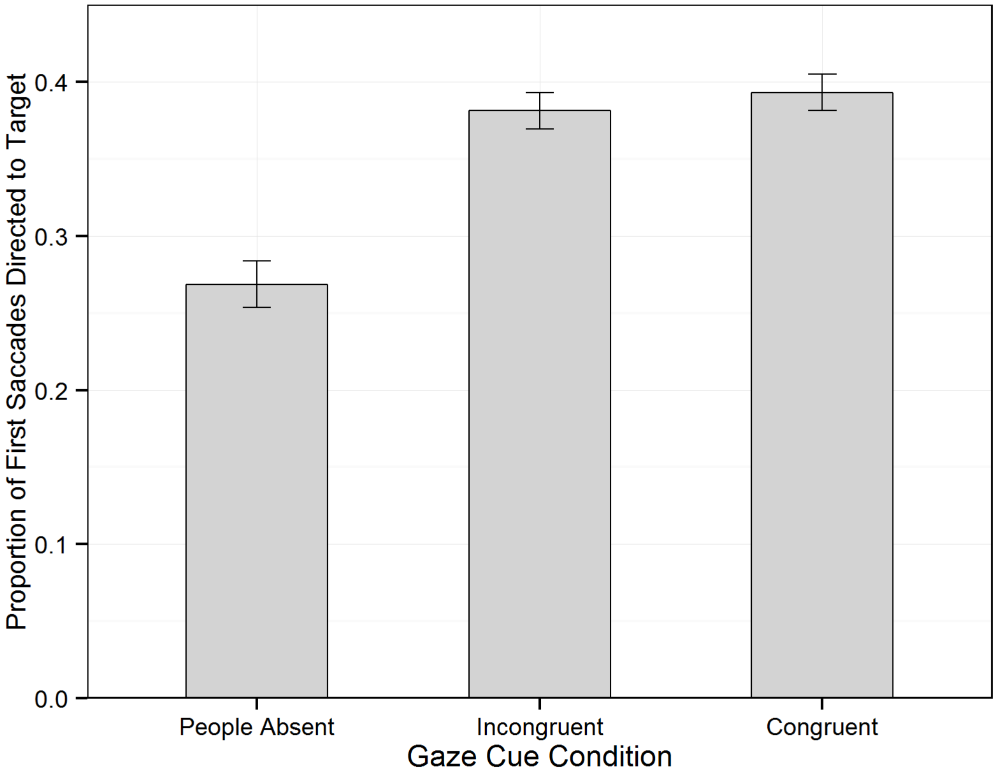

Figure 10. The proportion of first saccades directed toward the target across three gaze cue conditions. Error bars show standard error across all data samples.

Having a person present in the scene resulted in first saccades that brought the eyes closer to the target. The distance from the landing point of the first saccade to the centre of the target boundary box – that is, the end point accuracy of the first saccade – was significantly reduced in these conditions. Analysis of all trials showed that even unhelpful incongruent cues facilitated accuracy, with significantly shorter distances to the centre of the target in this condition than in the person absent condition $\beta = -1.458$, $SE = 0.229$, $t = -6.347$. Accuracy of the first saccade sees even greater improvement in the congruent gaze cue condition, where the gaze cue is directed towards the target object $\beta = -1.736$, $SE = 0.222$, $t = -7.795$. These results are displayed in Figure 11.

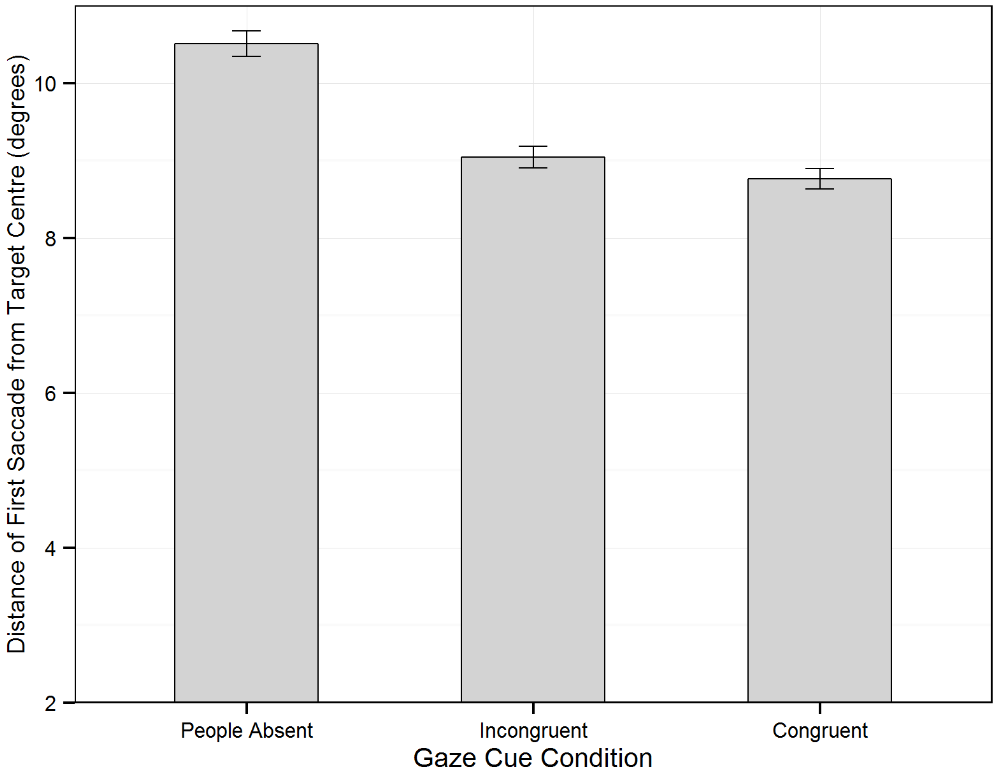

Figure 11. The distance of the landing point of the first saccade from the centre of the target ROI (in degrees of visual angle) as a measure of end point accuracy across three gaze cue conditions. Error bars show standard error across all data samples.

As in the other search initiation measures, follow up analysis found no significant difference in the accuracy of the first saccade between the incongruent and congruent gaze cue conditions, $\beta = 0.277$, $SE = 0.187$, $t = 1.476$.

Scene Scanning

After initiation of search, several measures were grouped together as a means of understanding scene scanning behaviour. First, the time taken to first fixate on the target was considered, which provides a measure of how quickly the observer successfully locates the target even if they do not press the trigger button at that time. However, manual response times were also measured, which provide a broader measure demonstrating how long participants took to locate the target and decide it was the correct object, indicating search completion by pressing a trigger button on the gamepad. Following from this, scan path ratio allows the determination of the overall efficiency of the search made by participants. This was calculated by taking the length of the actual search route (the sum of all saccade amplitudes) and dividing it by the optimal route (the distance from the starting point of search to the centre of the target). Each of these measures examined only 'correct' response trials; that is, trials in which participants fixated the target at some point.

Error rate was used to determine the success rate of participants, and as expected uses data from all trials. In this study, an error was defined as when the participant made a false-positive response; that is, pressing the trigger button on the gamepad when no fixation on the target object had occurred. As a measure of error it is possible that this misses some aspects of search and detection, namely that people can find and identify targets without fixating them (Henderson, McClure, Pierce & Schrock, 1997). The design of the study limits the extent to which this peripheral identification can be accounted for. Fixations that landed close to, but not within, the target boundary box may have allowed participants to identify the target, as is clear from Henderson et al.'s (1997) study that identification was very accurate without foveal vision. However, with eye tracking information being the only data gathered from search, it is impossible to tell whether object recognition occurred within these parafoveal fixations. Therefore, the strict criteria of search success is required – we can only be sure the participant has correctly located the target when they have both fixated it and pressed the gamepad button to indicate search has been concluded.

As a final point, the extent to which there was any overt orienting to, and selection of, gaze information within the search was investigated by considering the number of fixations on the person's face within the scene. This was again determined as a fixation which landed within the boundary box drawn around the person's head.

Having a person present in the scene was of great benefit for improving time taken to first fixate on the target object, with significant reductions for both person present gaze cue conditions. To satisfy model assumptions logarithmic transformation of the data was required, as was the removal of correlation parameters to achieve convergence. The facilitation effect was particularly strong in the incongruent gaze cue condition, $\beta = -0.101$, $SE = 0.013$, $t = -7.78$, which became even stronger when the cue was changed to a helpful, congruent cue, $\beta = -0.114$, $SE = 0.013$, $t = -8.29$. These results are displayed in Figure 12.

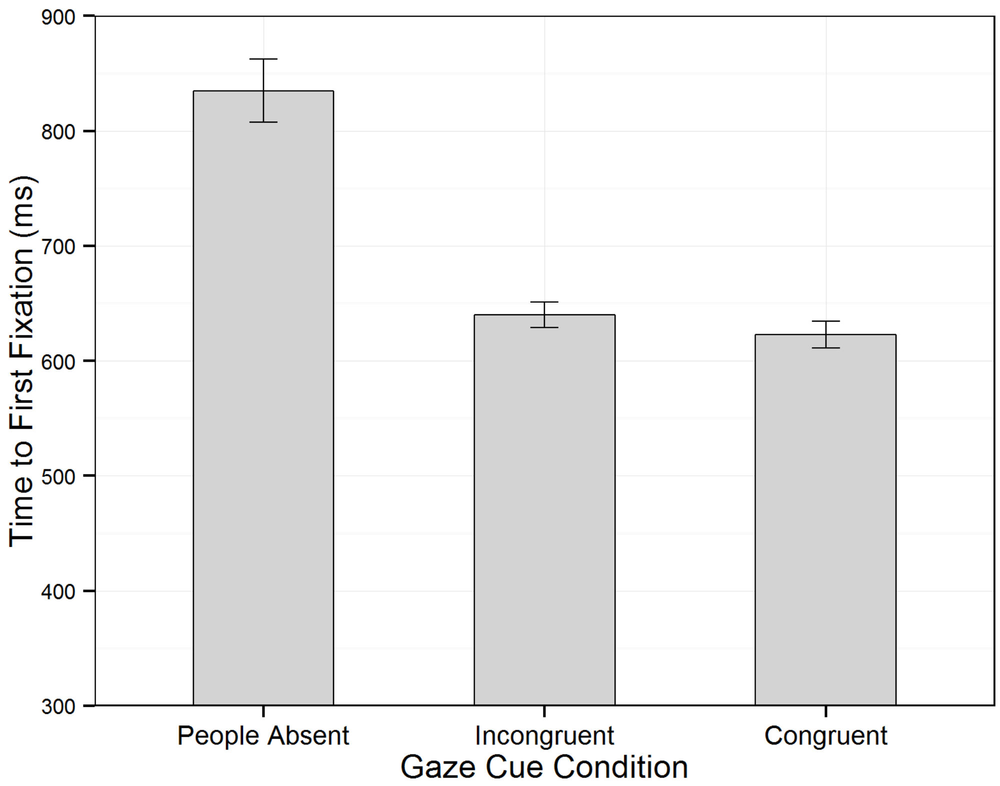

Figure 12. The time to first fixation on the target (ms) from scene presentation across three gaze cue conditions. Error bars show standard error across all data samples.

As might be expected from Figure 12 above, the difference between the incongruent and congruent gaze cue conditions was apparent, but did not reach significance $\beta = 0.012$, $SE = 0.009$, $t = 1.28$. Following a similar pattern to that described in search initiation, it seems that the simple presence of a person in the scene, regardless of the cue they are giving, greatly improves the speed with which participants fixate on the target for the first time. Although there is a smaller improvement when the cue changes from unhelpful to helpful, this does not produce strong enough changes in observer behaviour to elicit a significant effect.

As with the time to first fixate measure, response times required a logarithmic function to transform the data into a normal distribution. With this complete, the LMM showed again a significant effect of person presence within the scene. Both person present gaze cue conditions showed dramatically reduced response times in comparison to the person absent condition, with the congruent condition, $\beta = -0.128$, $SE = 0.011$, $t = -11.10$, having a marginally stronger effect than the incongruent condition, $\beta = -0.120$, $SE = 0.011$, $t = -10.87$. There was no significant difference between the two person present gaze cue conditions, $\beta = -0.007$, $SE = 0.007$, $t = 1.11$. These results are shown in Figure 13 below.

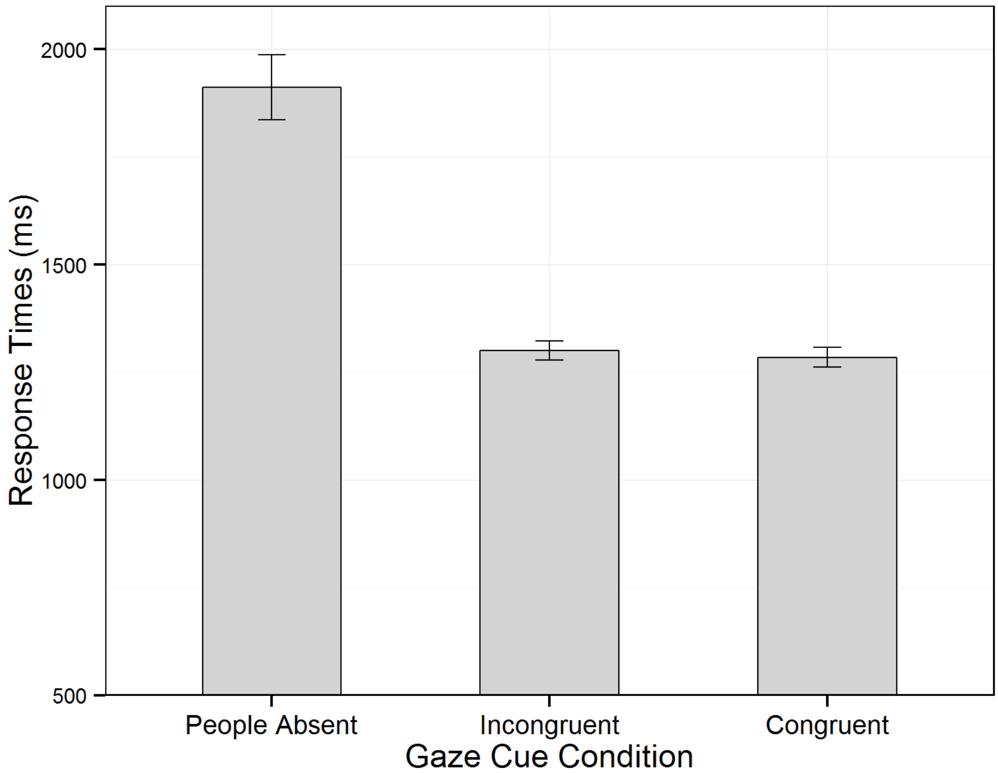

Figure 13. Response times (ms) to button press indicating successful search for the target across three gaze cue conditions. Error bars show standard error across all data samples.

Considering search efficiency, an optimal route would be a single saccade from the fixation point at scene onset to the centre of the target object; therefore more efficient searches should have a scan path ratio closer to 1. Although in this model the data were not normally distributed no transformation was performed as for search efficiency it is to be expected that the majority of responses fall at the lower end of the scale with a lower ratio. The results are presented below in Figure 14.

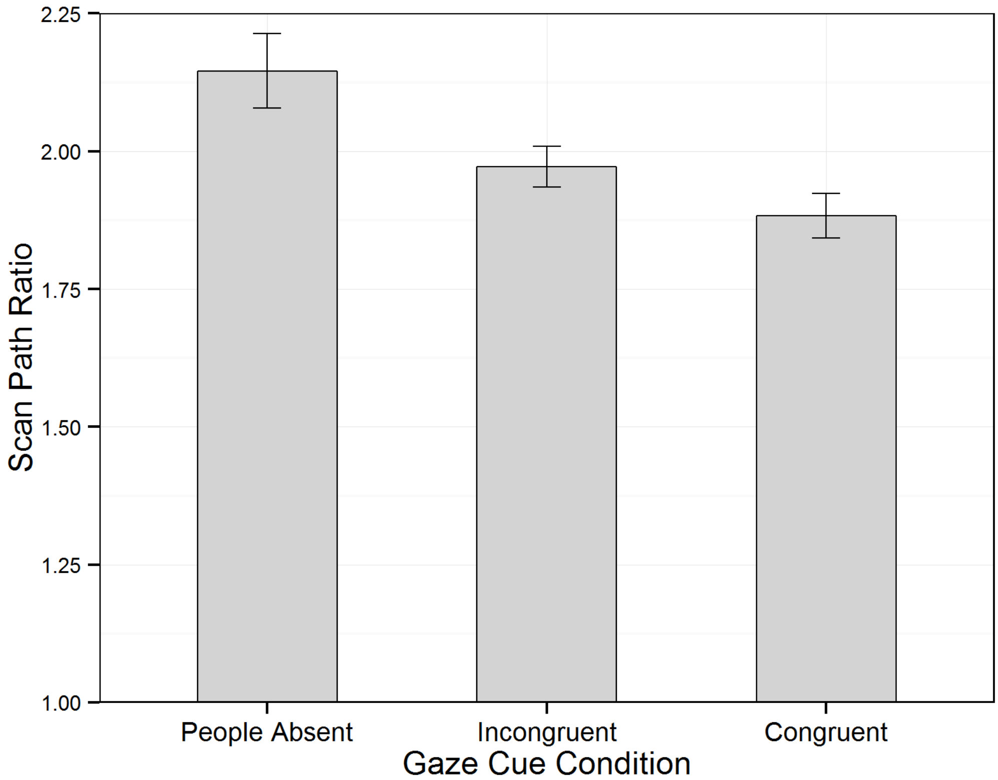

Figure 14. The scan path ratio across three gaze cue conditions. Error bars show the standard error across all data samples.

Analysis showed that there was a significant improvement in search efficiency when a person was present in the scene, for both incongruent, $\beta = -0.163$, $SE = 0.077$, $t = -2.111$, and congruent gaze cue conditions, $\beta = -0.025$, $SE = 0.077$, $t = -3.325$. While there is some improvement in search efficiency in the congruent gaze cue condition as compared to the incongruent condition, $\beta = 0.093$, $SE = 0.054$, $t = 1.72$, this did not reach significance.

With a definition of error as the proportion of trials in which participants made false-positive responses, the error rate is already very low in baseline person absent trials. However, as can be seen in Figure 15, when a person is introduced to the scene the proportion of errors drops to almost negligible levels.

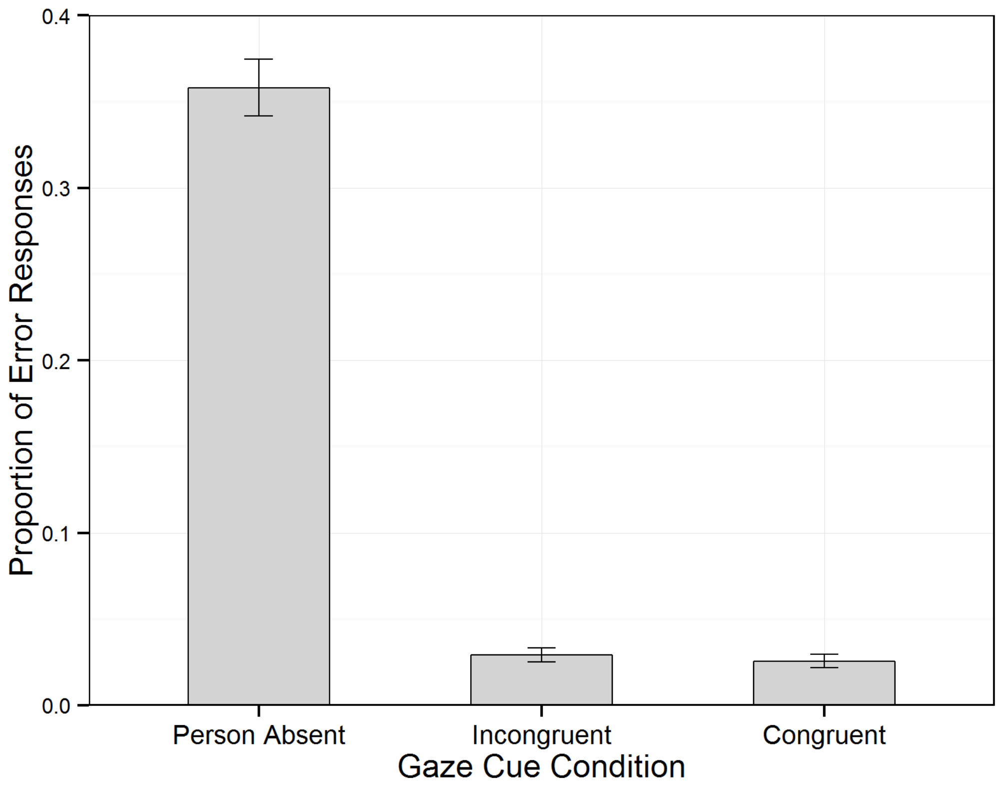

Figure 15. The proportion of false-positive responses in all trials across three gaze cue conditions. Error bars show the standard error across all data samples.

When compared to the person absent conditions both the incongruent, β = -0.329, *SE* = 0.018, *t* = -17.58, and congruent gaze cue conditions, β = -0.332, *SE* = 0.019, *t* = -17.29, resulted in drastically smaller proportions of errors made. As would be expected from Figure 15, error rate in the two person present gaze cue conditions was almost identical ($t < 0.5$).

Finally, the extent to which overt fixations on the person in the scene occurred was considered using ‘correct’ trials; trials where no false-positive response occurred. This is simply because in trials where participants make an error response, it is impossible to be sure of what type of strategy they were using for the search or what might have caused them to make a false-positive response. To measure the extent of overt gaze seeking, the total number of looks towards the person in the scene was calculated, with a ‘look’ defined as a fixation on the head of the person.

Analysis showed that across a total of 3345 trials where a person was present, only 157 involved a fixation on the person's face, which accounts for just 4.69% of trials.

Discussion

This study examined the effect of gaze cues provided in photographs of real world scenes on observer eye movement behaviour when searching for targets. Gaze cues were provided by a person sitting behind the table upon which the objects were arranged. Having a person present in the scene benefits both search initiation and scene scanning, but there were no apparent effects of gaze cue congruence with target location. These results allow us to quantify the benefits of gaze cues on visual search in a more realistic context than traditional Posner-type paradigms.

Within the search initiation phase there are clear effects of person presence on eye movement behaviour. The time taken for pre-saccadic launch processing is relatively similar across all three gaze cue conditions, but once a gaze cue is directed toward the target a significant reduction in the time taken to launch the first saccade is seen. First saccade latency was slightly reduced in the incongruent gaze cue condition, so there does seem to be some evidence that person presence facilitates pre-launch processing, but that in the incongruent condition this benefit is not strong enough to elicit a significant effect. Faster first saccade latencies as a result of person presence suggest that participants spend less time processing information before launching their first saccade. The very small proportion of trials in which the person's face was fixated rules out overt gaze following as a means by which person presence could facilitate pre-saccade launch processing. However, the role of covert attention could be considered as an alternative possible explanation.

Intuitively one would assume that extra information to process – in this case, the gaze cue provided by the person in the scene – would require more processing time and should in fact result in longer first saccade latencies. This is not reflected in the results described above. Carrasco and McElree (2001) may provide an explanation as to why this is the case. They cite the importance of covert attention in processing visual information, where it can be used as something of a filter: we can covertly attend information at a cued location without making an eye movement, thus granting this information priority in visual processing. The main issue Carrasco and McElree (2001) focused on was whether covert attention also allows speedier processing of visual information. They used the response-signal speed-accuracy trade-off (SAT) procedure in order to measure both discriminability of the stimuli and the speed of information processing in two visual search tasks. In one task participants performed a feature search where they looked for a target in an orientation different from the orientation of the distractors around it. In the other task participants performed a conjunction search, which required searching for a unique combination of two features. Their results showed that covert attention improved both the discriminability of features and the speed of information processing. Carrasco and McElree (2001) were able to explicitly show that spatial cues resulted in faster processing in both search tasks. The authors state that their results are consistent with the theory that parallel processing occurs across all locations in a scene, but that spatial cues permit attention to be focused in a restricted region. In terms of the current study, person presence may encourage restriction of processing to the lower half of the scene containing the table, or even to one side of the table. With covert attention deployed to this area, processing there is given priority resulting in less time needed to launch the first saccade.

Once the saccade has been launched, strong effects of person presence become apparent, regardless of the type of gaze cue presented. In measures of both first saccade direction and end point accuracy of the first saccade, performance was much better when a person was present in the scene, resulting in a higher proportion of first saccades directed toward the target that were also more accurate, bringing the eyes closer to the centre of the target object. Considering first saccade direction, it is helpful to first establish the chance rate of a first saccade being directed toward the target that might be expected if participants were deploying their first saccades randomly. Theoretically, the target object could appear anywhere within the scene, offering participants a 360° radius in which to deploy their saccade. The target object is defined by a 45° cone (22.5° above and below the centre of the target), which would result in the first saccade being directed toward the target in 12.5% of trials by chance. In all gaze cue conditions the proportion of first saccades directed toward the target is higher than this chance percentage, so it can be assumed that there is some guidance of saccade direction at work. If, as Carrasco and McElree (2001) suggest, spatial cues allow the restriction of search space attended, the presence of a person would at the very least restrict search to the bottom half of the scene where the table is present. This reduces the search radius to 180° and the likelihood of directing the first saccade to the target by chance to 25%. Again, the proportion of first saccades directed to the target exceeds this percentage, suggesting participants' first saccades are not simply allocated by chance.

If restriction of the search area is the sole cause for improvement in first saccade direction when a person is present, it would be expected that the proportion of first saccades directed toward the target in person present trials should be twice that of the proportion found in person absent trials since only half of the scene is being searched. In the current study, the proportion of trials in which the first saccade is directed toward the target increases from 26.8% in person absent trials to 38.1-39.3% in person present trials. Although this is slightly less than double the person absent proportion, the increase is great enough that it is possible to safely assume restricted search, guided by person presence, plays a large role in improving accuracy of the first saccade.

The measure of end point accuracy of the first saccade – the other main measure of saccade accuracy – provides further support for a search area restriction hypothesis. In this measure the distance from the centre of the target to the landing point of the first saccade is always quite large (in excess of eight degrees of visual angle), which would suggest that the first saccade does not really bring the eyes that close to the target. However, person present gaze cue conditions result in significantly more accurate first saccades than the person absent gaze cue condition. Therefore, while there is still quite a bit of distance to be covered, person presence brings the eyes slightly closer to the target because search space is restricted.

How then does person presence assist in reducing the search area, improving all measures of search discussed in the current study? Viviani and Swensson (1982) performed a study which may begin to explain the role of person presence facilitating search in the current study. They required participants to identify a target located in the periphery of their visual field amongst an array of distractor items, and to fixate it as quickly as possible in one eye movement. The further the target was into the periphery, the greater the saccade latency, proportion of movement errors, and the proportion of first saccades followed by a corrective eye movement. In their second experiment, Viviani and Swensson (1982) informed participants how far from centre the target would be, which saw performance efficiency return. It is possible that person presence in the current study may act somewhat like the clarification given to participants in Viviani and Swensson's (1982) study, allowing them to make more accurate eye movements to the target.

Horowitz and Wolfe (2001) agree that the processes that control the time taken to find a target are more complicated than simply locating it. Instead, these processes must incorporate memory. A template of what the target may look like is held in working memory so that when a possible match for the target is located a verification process begins. To verify the potential candidate as the actual target, the template in short term memory must be matched with the target's known identity in long term memory. This more lengthy process is a good reason to include both measures of search speed in the scene scanning section of the results – first fixation time illustrates the point at which the participant identifies the potential target, with response time providing a measure of the added time required for target verification.

Without this more sensitive measure, an important finding may have been missed. As can be seen from Figures 12 and 13 in the results section, the time taken to first fixate on the target is approximately half of the overall response time measured in all three gaze cue conditions. This means that participants are fixating the target relatively quickly (within ~640ms for person present gaze cue conditions and ~1900ms for the person absent condition) but taking the same time again to press the trigger button, perhaps using this time to conduct the verification procedure discussed by Horowitz and Wolfe (2001). Knapp and Abrams (2012) have shown that verbal cues take much longer to process than pictorial cues and that this can affect visual search even when ample time is allowed to process the verbal cue. It could be assumed then that difference from time to first fixation and the button press is a result of identifying a visual target from a verbal name requiring extra time during verification to match the observed object to the template in memory.

Schmidt and Zelinsky (2009) expand on how this target template may be accessed from verbal rather than pictorial cues. As has been discussed previously in this chapter, Schmidt and Zelinsky (2009) argue that in order to make laboratory-based research mimic real world search more closely, we must provide stimuli that match more consistently with what is experienced in the real world. To that end, they gave participants a visual search task with categorical cues (i.e. a verbal description like "*teddy bear*") which had varying degrees of specificity. Compared to a control condition that used a pictorial cue, categorical cues were either an abstract textual description of the target; a precise textual description; abstract and colour textual description; or a precise and colour textual description. The authors found that as the specificity of the target information increased and that this occurred in a gradual way, rather than an all-or-none cuing effect. Although abstract textual descriptions offered the least amount of guidance, there was still some guidance present. A precise textual description further improved search guidance, though there does seem to be an upper limit to the degree to which textual cues can guide search – none guided search as accurately as a pictorial cue. This was supported by further research conducted by Yang and Zelinsky (2009). They also asked participants to complete a visual search task in which they were given either a target preview or a categorical cue, with realistic distractor items randomly included in each trial. While participants made more eye movements in the categorical cue condition and their overall search time was longer than the target preview condition, targets were still fixated much earlier than would be predicted by chance. Together, these studies suggest that categorical cues, which were given in the current study, do facilitate search by permitting early access to the target template. In the current study, the target cue appears for a full 500 ms, followed by a 500 ms blank screen presentation, before the search scene appears. This allows considerable time in which the target template can be accessed from memory, guiding search through the array of real world objects.

Malcolm and Henderson (2010) suggest person presence may offer further benefits in using the target template due to how it can assist observers in disregarding distractors. Conducting search through real world scenes means that arrays are relatively large. The authors conducted a visual search task using real world scenes with two different cue types (verbal and pictorial), and manipulated whether the target object appeared in a likely or unlikely position within the scene. They measured fixation placements across the scenes to determine what processes participants used to guide their search. Their results led them to the conclusion that during scene scanning the visual system combines all possible sources of top-down information to decide fixation placement and duration. To put this into context, Malcolm and Henderson (2010) discuss how the scene's context (e.g. office area, kitchen) provides global information allowing the observer to identify areas where the target is likely to appear. Simultaneously the target template allows the distinguishing of possible targets from distractor items. In the current study, the global context of the scenes remains unchanged across all gaze cue conditions, but it also prevents participants identifying any likely area of the scene where the target will appear other than the table top, which does not rule out any distractors. Therefore in person absent scenes the participants have no further information to guide their eye movements, and so must scan through the entire array to find a potential match to the target template. However, when a person is in the scene, they highlight areas where the target may be likely to appear, thus adding to the global information available and reducing the number of potential targets within the array. This also serves as a logical explanation for the reduced rate in errors that occur when a person is present in the scene. The inclusion of more global information makes it much less likely that a participant will end a search without fixating the target.

In all measures presented in this chapter, there were clear benefits of person presence, but none of congruency. Intuitively, one would expect the greatest benefit to come from a gaze cue that is helpful. In the Posner-type studies discussed in the introduction, there were clear facilitation effects when the cue given was congruent with target location (e.g. Friesen & Kingstone, 1998; Ricciardelli et al., 2002) Gaze seeking and following was also apparent in more realistic studies. For example, Castelhano, Wieth and Henderson (2007) gave participants a series of photographs that told a story. When an actor was present in a photograph the actor’s face was very likely to be fixated, as were objects that were cued by the gaze of the actor.

It is possible that the differences in complexity of stimuli may account for some of this discrepancy in facilitation effects. Some of the earliest research into scene viewing describes the patterns of fixations across the scene as dependent on scene content (Buswell, 1935). Following this line of thought, an explanation for the discrepancies in expected congruency effects may come from research by Downing, Dodds and Bray (2004). While recognising the large volume of research that defines gaze as a special type of stimulus, the authors wondered whether if, in certain contexts, gaze attracted visual attention simply because of its spatial compatibility between the gaze cue and the target. To test this hypothesis, they conducted a study which followed basic Posner-type principles of a centrally presented face followed by a directional cue and a letter identification task. Downing et al. (2004) used two different cues. One was the traditional gaze cue, beginning with eyes closed then after 907ms shown open and pointing either to the left or right. The second cue followed the same procedure, but instead of the eyes opening, the mouth opened and the tongue was pointed to either the left or right. Results showed that when neither cue was predictive of target location, there were identical attentional shifts following a gaze or tongue cue. The authors concluded that while gaze certainly holds status as a special stimulus, this is not best exhibited by orienting paradigms; instead more complex properties of gaze were required to truly study its specialised nature. The key point here is that when cues were non-predictive, as they are in the current study, a tongue cue was as effective as a gaze cue simply because of its spatial compatibility with the target. In terms of the current study, this would suggest that both incongruent and congruent gaze cues are equally effective in facilitating search because they have the same spatial compatibility with the target.

In addition to these specific spatial compatibility effects, it is possible there is an explanation that accounts for the lack of congruency effects in considering the broader social implications of person present scenes. Social facilitation, commonly understood as the effects humans have on one another, particularly on an individual basis, may begin to have an effect on participants' performance as soon as a person is introduced to the visual search scene (Guerin, 2010). Social facilitation was originally explored by Triplett (1898) who examined how cyclists' performance changed depending on what type of race they participated in. He found that cyclists' performance improved when they raced alongside other people compared to when they cycled alone. This research was the basis for a large volume of research in social psychology, perhaps the best known of which is by Allport (1920), who coined the term ‘social facilitation). Allport (1920) asked participants to work on various tasks including vowel cancellation, problem solving, and multiplication, either alone in a cubicle or alongside others at a common table. With the exception of problem solving and judgement tasks (where participants had to judge odours and weights of objects), participants performed better when working at a common table than they did when working alone in a cubicle. In a later review Zajonc (1965) attempted to integrate the results from a number of studies that had followed Triplett (1898) and Allport’s (1920) efforts, whose results presented varying levels of support for the original hypotheses of social facilitation. As a result of his review, Zajonc (1965) proposed that essentially Allport (1920) was correct: having other people around you doing the same task makes you perform better. However, Zajonc (1965) suggested that not only did this facilitation occur in motor tasks, as Allport (1920) had suggested, but in all tasks. Considering this evidence in relation to the results of the current study, it is quite possible that social facilitation may account to some degree for improved performance in person-present scenes compared to person-absent scenes. If the mere presence of a person is enough to improve one’s performance in a task, it is plausible that the presence of a person in the visual search scene may provide the same level of social facilitation as described by Triplett (1898), Allport (1920), and Zajonc (1965). Interpreting the results in this way would explain why facilitation occurs in person-present scenes without any gaze cue congruency effects: it is not the gaze cue itself that is being utilised to enhance performance, but the additional stimulus of the person – as a whole – in the scene.

As a final consideration in the data analysis, the results showed very little evidence of overt gaze seeking. This contradicts a mass of evidence that suggest when a person is present in the scene they will be preferentially fixated (e.g. Birmingham et al., 2007, 2008b; Fletcher-Watson et al., 2008). It is possible that Downing et al.'s (2004) explanation of gaze as special during orienting tasks only because of its spatial proximity to the target goes some way to account for this. In other words, the type of task assigned modifies the type of stimuli that is useful. Once a scene begins to feature more complex content, there are more cognitive processes involved. Participants are no longer simply determining if a target appears or not, or what that target is, but are searching for it within an array of other objects. However, other more complex cognitive processes may also play a role, particularly task demand. Adaptations in eye movements resulting from changing task demands were found by Itier, Villate and Ryan (2007). They identified that while it was accepted throughout social attention research that eyes – and gaze – are central to processes within this field, the extent to which eyes attract different levels of attention based on task demand was relatively unexplored. They set out to determine the level of attention allocated to gaze in two different tasks by tracking eye movements. In one task, participants were asked to determine the gaze direction of a person within the presented scene; a task which required specific fixation on the face, and particularly the eye region. In the other task participants had to instead determine head orientation. Itier et al. (2007) found evidence of gaze processing in both tasks, shown by a deterioration in performance (RT and accuracy) when the head orientation and gaze direction were not congruent with each other. Perhaps the most critical finding, in terms of explaining the results of the current study, was that while the eye region was the end point of approximately 90% of first saccades in the gaze direction task, only around 50% of first saccades landed in the eye region during the head orientation task. These results are also inconsistent with a reflexive-orienting model of social attention, where eyes will automatically fixate on a person within a scene. Instead, Itier et al. (2007) argue the results demonstrate an endogenous, top-down method of allocating attention.

In the current study participants were given no specific direction with regards to the person in a scene, only that they were to find the target object as quickly as possible. In order to meet the demands of the task – focusing on speed – it makes sense that participants would rapidly allocate attention to the location where that target would be found: the table top. Furthermore, Itier et al. (2007) found only around 50% of fixations on the eye region within stimuli that only presented a portrait representation of a person (head and shoulders). Perhaps then, in the current study where the face is presented alongside all other natural cues of a person (i.e. a body), it can be expected that the proportion of fixations on the eye region may drop further. In tasks where a head and shoulders is presented centrally within a screen, they take up a large proportion of the viewing angle. Even though typical Posner-type studies inform participants the cues provided will be unhelpful, their central presentation, in addition to observers' fixation at the start of a trial already occurring on the face, means that it is hard for observers to ignore the gaze cue. As such, reflexive orienting as found in these studies is perhaps unsurprising. When the cue is given within a natural context, as in the current study, participants are not fixating on the head area when the scene appears. To overtly seek out gaze would require an eye movement. Furthermore, in the current study the head – and in particular the eye region – of the person in the scene take up a much smaller proportion of the visual field. Again, this could lead to further, expected, drops in fixation rate on the person within the scene.

The current study presents elements of Posner-type tasks, namely a central cue to a target in the periphery, in a more realistic context by using photographs of real scenes containing an individual giving a gaze cue that involved both head and eye movement, rather than keeping the head stationary. Results showed that the presence of a person in the scene facilitated performance across all aspects of search from search initiation to scene scanning. However, there were no apparent effects of gaze cue congruency with target location, which seems counter-intuitive. A possible explanation for this comes from Downing et al. (2004) who suggest that when cues are non-predictive, as is the case in the current study, a gaze cue is useful just because of its spatial compatibility with the target regardless of its congruency. Alternatively, it may be that the presence of a person in the scene provides social facilitation (e.g. Allport, 1920; Zajonc, 1965), thus improving performance regardless of the gaze cue provided. There was also little evidence of overt gaze following. However, since there are still clear benefits of person presence, it can be concluded that participants' search is still guided to some extent by person presence. It is possible that person presence helps to reduce the search area in which attention is deployed, by adding to the global context of the scene. This, in collaboration with the target template accessed from memory following the textual cue, results in improved search performance across a number of measures when a person is present in the scene. As discussed at the beginning of the chapter, the current study was designed to provide a baseline for further research. For that reason, participants were given no instruction regarding the purpose of the presence or absence of a person within the scene. However, most Posner-type paradigms give some mention of the face cue in the task, even if it is just that the cues provided by the face will not be useful for completing the task. Therefore, the logical next step would be to utilise this paradigm in an exploration of how instructions concerning the presence of a person in the scene affect visual search behaviour. This will be investigated in the next chapter.

Chapter Three

How does manipulating perceived helpfulness of gaze cues affect eye movement behaviour in a visual search task?

Introduction

The previous chapter explored a novel and more realistic Posner-type paradigm and how the use of this methodology might support or challenge findings of previous research that used more simple stimuli. It was found that when observers were presented with scenes that featured a real environment, with a real person giving a natural gaze cue that included head and eye movement, person presence facilitated participant performance across all measures of search. From search initiation and throughout scene scanning, participants were more accurate, faster and more efficient when a person was present in the scene. The study discussed in the previous chapter provides a baseline from which the new paradigm can be developed and added to. One aspect of the original Posner-type studies was excluded from the study discussed in Chapter Two because it deserved a thorough investigation of its own: the instructions given to participants about their task. Task instructions provide a means by which participants' beliefs about the task can be manipulated, changing what is made salient to them or shifting their perception of the focus of the task. This chapter explores how two different types of instruction affect participants' performance when searching in scenes containing social gaze cues.

In studies that have used modified versions of the Posner task to study how we orient to social gaze cues, the instructions given to participants vary. In some studies, participants are explicitly told to ignore the gaze cues provided (Ricciardelli, Betta, Pruner & Turatto, 2009; Ricciardelli et al., 2002). Others instruct participants that the cues would not be helpful for the task (Friesen & Kingstone, 1998, 2003a), or that the face is simply there to provide a fixation point to return to at the end of each trial (Driver et al., 2009). These studies use similar paradigms, but provide different information to participants about the purpose of the face in the scene. Despite this, all are in agreement concerning the evidence they produce: regardless of what participants are told about the purpose (or lack thereof) of a face in the scenes shown to them, the same reflexive orienting responses are found. Even in conditions where the gaze cues are non-predictive of target location – including experiments where participants are informed of this – the reflexive orienting response to the cued location occurs.

However, in tasks that use less constrained paradigms the effects of varying instructions are quite different from those cited above. A classic example of the power of instruction comes from Yarbus (1967) who used the painting ‘An Unexpected Visitor’ by I. P. Repin to study how observers viewed the scene. He asked the same individual to view the painting several times, each time with a different task. On one occasion they freely viewed the painting, on another, they were asked to gauge how long the visitor had been away, the material wealth of the family, to memorise the content of the scene, or to estimate each person’s age. Depending on the task given, the way in which the observer examined the scene changed. When asked to rate the wealth of the family, the observer fixated on clothes and furniture, but when freely viewing the painting fixations clustered around the peoples' faces. Yarbus' (1967) study provides a clear example of how eye movements can change when viewing the same image, solely as a result of instructions given to the observer.

Yarbus' (1967) findings became the foundation for a growing body of research that investigated this link between eye movements and the complex cognitive processes involved in visual tasks. Developments in eye tracking technology, particularly the advances in mobile eye tracking equipment, permit an exploration of eye movements in much more natural task settings, which Hayhoe and Ballard (2005) discuss as being particularly useful for exploring how task instructions impact gaze behaviour. For example, while asking someone to make a cup of tea will produce a similar pattern of eye movements to people who are asked to make a sandwich, the focus of these fixations vary depending on the task assigned (Land & Hayhoe, 2001). This study, and those that explore similar natural everyday tasks, again produce one salient point upon which they all agree: fixations are very closely linked in time to the evolution of task demands. Fixations do not tend to occur where visually salient stimuli are present; instead they operate in a 'step ahead' format. A fixation will be targeted to the task-relevant object – for example, a mug or cup during the task of making tea – prior to any manipulation of that object, but the next eye movement will be programmed and executed before manipulation of the object has finished. This means that they eyes stay a step ahead of the hands, moving to the next stage of the task and fixating on the relevant object before any actual interaction with it occurs (Land & Hayhoe, 2001). These fixations are deployed with the purpose of obtaining specific information that relates to the task, and this pattern of eye movements has been referred to as 'visual routine' (Hayhoe & Ballard, 2005).

The question then is how we can accurately explore the extent to which slight manipulation of task demands (as defined by the instructions given to participants) affect eye movement behaviour. It is possible that experiments conducted in the real world where real interaction occurs between the participant and their task produces greater variation in the impact of task on eye movements than in experiments conducted in the laboratory. The evidence presented by Land and Hayhoe (2001) and discussed by Hayhoe and Ballard (2005) demonstrates that task instructions can change the way participants move their eyes around an environment, which has not been found in the more controlled laboratory settings of Posner-type tasks. What then might elicit a change in eye movement behaviour in response to task instructions within this more controlled setting? The experiments presented in this chapter follow the same procedure as the study discussed in Chapter Two: participants were asked to find the target object as quickly as possible, creating an emphasis on speed. It is possible that by adding a second instruction regarding person presence, a competing task demand is created. Schwartz et al. (2005) used functional magnetic resonance imaging (fMRI) to compare the effects of competing endogenous and exogenous attentional effects, and their interaction in the brain. Participants performed a visual detection task where they were required to monitor a letter stream for a pre-specified target within a rapid successive visual presentation of letters that presented one letter every 750ms. This letter stream was presented continuously either alone or accompanied by checkerboards that flickered in the periphery to the left or right of the letter stream, or in both areas. Participants were assigned either to the high-load condition where they had to find an infrequently appearing letter or a low-load condition where they had to look for a specified colour. Schwartz et al. (2005) found that in the high-load condition participants allocated less attention to the salient flickering peripheral stimuli than those in the low-load condition. Furthermore, the activation in various regions of the visual cortex was decreased for these peripheral stimuli when participants were experiencing high attentional load. These results show that participants will allocate attention to the critical element of the task. In the current study, there is an initial emphasis placed on speed, but by introducing person presence to the instruction, it is possible this creates the same competing demand as the peripheral stimuli in Schwartz et al.'s (2005) study.

It is clear that the impact of instruction on eye movement behaviour varies considerably across different tasks. The controlled settings of Posner-type tasks seem to produce very similar responses regardless of what participants are told, with each study eliciting reflexive orienting responses to non-predictive gaze cues. However, when eye movements are explored in the real world, the fixations of participants vary considerably depending on the task they are given. The first example of this comes from Yarbus' (1967) seminal study where participants viewed the same painting with different tasks on each viewing, and this has informed later research which has been able to employ more sophisticated means of tracking eye movements. These later studies (e.g. Land & Hayhoe, 2001) show that across different tasks the differences in eye movements can be subtle, following similar patterns (i.e. the eyes staying a step ahead of the hands) but responding to the individual demands of the task. The paradigm used in the experiments discussed within this chapter incorporates some elements of Posner's (1980) paradigm, in that it has a central face that provides gaze cues to one or other side of the scene, but otherwise is quite different. It uses a real person performing a natural gaze cue that incorporates head movement in addition to movement of the eyes, and that person is situated within a real environment. The task is also more complex, involving searching for a target within a large array. In this more demanding task, it is possible that manipulating the instructions given to participants may elicit more varied eye movement responses than those seen in previous laboratory-based studies. This chapter explores how instructions specific to the purpose of the person within the scene impact on eye movements during the visual search task. When added to the original instruction that emphasises speed (*"Please find the target as fast as you can"*), evidence from Schwartz et al. (2005) suggests the person-specific instructions may create a competing task demand. To allow comparison to previous Posner-type task studies, two different instructions regarding person presence were used, and are designed to manipulate the perceived helpfulness of the gaze cue offered within person-present scenes. The first condition replicates the most common instruction given in the previously cited literature, where participants are told to ignore the presence of the person in the scene – it is suggested their cue is unhelpful. The second instruction condition suggests that the person in the scene may be helpful in finding the target object. Across both instruction conditions the gaze cues provided are non-predictive, and so any difference in eye movement behaviour can be attributed solely to the effect of instruction.

Method

Participants

A total for 40 people (16 male) were recruited for the unhelpful instruction condition, and a further 40 people (14 male) were recruited for the helpful instruction condition. All had normal or corrected vision and were naïve to the purposes of the study. Level one and two undergraduate students received course credits for participation; anyone not eligible for course credit was paid £2. Individuals recruited for participation in these studies had not participated in the previous experiment discussed in Chapter Two.

Materials

The materials used were the same as described in Chapter Two. However, to give a quick review, experimental scenes were created using ten different sets of everyday objects. Each scene featured one of the ten sets of 15 everyday items arranged on a table top. Within each scene one item was designated the target and another designated the distracter. These items were always on the opposite sides of scene centre, preventing any central bias (as discussed in Tatler, 2007). The target was equally likely to appear on the left or right side of the table. Every arrangement was photographed twice: once with the person in the scene looking toward the target and once with them looking toward the distracter, an example of which is shown in Figure 16.

Figure 16. An arrangement of an experimental scene showing the two photographs for one arrangement of objects on the table top. In this arrangement the target is the

Filofax and the distracter object is the earmuffs. Panel A shows the individual looking toward the target and Panel B shows the individual looking toward the distracter.

In Figure 16, Panel A provides an example of the *congruent* gaze cue condition where the person cues the target object. Panel B shows an example of an *incongruent* gaze cue condition where the person cues the distracter. Two versions of the experiment were created so that in version one, participants only saw Panel A of the arrangement in Figure 16, and in the other they only saw Panel B. In addition to the four person present arrangements for each object a final arrangement was photographed without an individual present to create a person absent control scene. In total ten person absent scenes and 80 person present scenes were created.

Eye Tracking

As in the previous chapter, participants' eye movements were tracked using an SR Research EyeLink 1000 eye tracker with a sampling rate of 1000 Hz, using pupil tracking and corneal reflection. The tracker was desk-mounted, sitting below the computer monitor and used to track a participant's dominant eye. The participant's head was kept stable throughout the experiments using an adjustable chin and forehead rest. Stimuli were presented on a 19-inch CRT computer monitor with a resolution of 1024 x 768 pixels. The Experiment Builder software developed by SR Research was used to run the experiment. Calibrations performed using the EyeLink 1000 were accepted if the average spatial error was less than 0.5 degrees and the maximum error was less than 1 degree over the 9 calibration points.

Procedure

The procedure used in these experiments was the same as described in Chapter Two, with one addition. The instructions given to participants, beyond the basics of how to perform the task, were designed to manipulate the perceived helpfulness of the gaze cue provided. In the unhelpful instruction condition, participants were told to ignore the presence or absence of a person in the scene, similar to previous Posner-type tasks, with the instruction: "*Some of the scenes will have a person in them, but please just ignore them. I'm using the same images over several experiments, but in this experiment the person isn't relevant; I'm only interested in how you search for the target object in the scene.*" Conversely, in the helpful instruction condition participants were told about the purpose of the person in the scene with the instruction: "*Some of the scenes will have a person in them. This person might be looking at the target, so they may help you find it faster.*" This instruction does not tell participants that they must look at the person; it simply provides them with more information about the context of the scene.

Data Analysis

Data were analysed for these experiments following the same procedure as described in Chapter Two using the R statistical analysis environment (R Development Core Team, 2011). To reiterate the salient points, lmer() functions return *z*- and estimated *p*-values for logistic models and *t*-values for linear models, within which we consider any effects for which the *t*-value is greater than two as reflecting a significant effect (as in Kleigl et al., 2012). In all models the random factors of participant and scene were included, and where possible the maximal model was used in which intercepts and slopes for the fixed effect of gaze cue condition was allowed to vary over both of the random factors (see Bates et al., 2014). Maximal models often fail to converge when large amounts of data are unavailable so random effects structures were simplified when necessary. In the analyses below the most complicated random effects structure that converged is reported.

Two different stages of search are reported in the analyses: search initiation (first saccade latency, first saccade direction, first saccade end point accuracy), and scene scanning (time to first fixate target, scan path ratio). Overall search behaviour is analysed in terms of the response times of participants to press the button, terminating search. The search initiation measures use data from all trials in the analysis irrespective of whether the participant later fixated the target, but measures of time to first fixation, response time, and scan path ratio only analyse trials in which the target has been fixated. Behaviour with respect to the individual pictured in the scenes (number of looks at the person) is also considered. These stages of search are explored within both instruction conditions. However, it should be noted that in the results section below the two instruction conditions will be considered separately; the analyses consider effects of gaze cue type within each instruction type.

Results

Unhelpful Instruction

Search Initiation

The first measure of search initiation to be considered is the time taken to launch the first saccade after the appearance of the scene (first saccade latency). For analysis, first saccade latency required two transformations. The data presented a small number of very short latencies, which most likely were the results of pre-emptive eye movements beginning before the appearance of the scene. Very short latencies were defined as being less than 100 ms. A total of 560 trials featured a very short latency (14.0% of the total number of trials) and these very short latencies were removed. These data are shown below in Figure 17, and are plotted prior to logarithmic transformation.

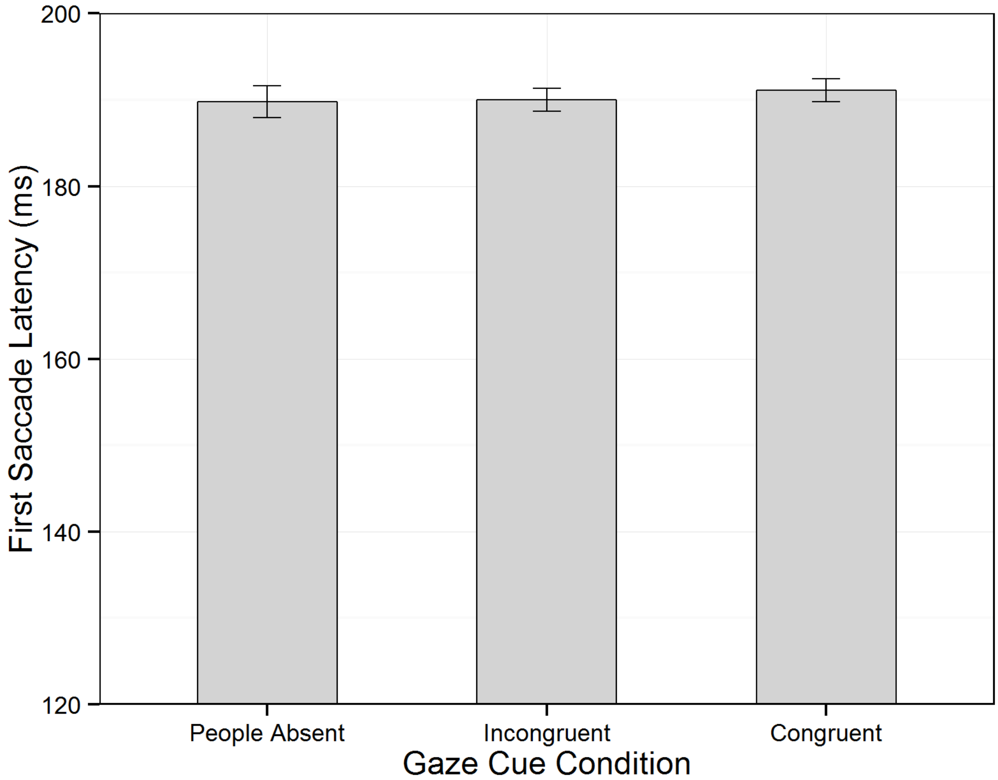

Figure 17. First saccade latency in each of the three gaze cue conditions. Error bars show standard error across all data samples.

In the first LMM, where the two person present conditions were compared to the person absent condition, the LMM did not converge and so correlation parameters were removed. There were no differences between the person absent and person present conditions, with the incongruent condition, $\beta = 0.020$, $SE = 0.014$, $t = 1.39$, showing numerically but not significantly longer latencies than the person absent condition, and first saccade latencies that were almost identical between the person absent and congruent conditions ($t < 1$). Follow up analysis showed no difference in first saccade latency between the incongruent and congruent gaze cue conditions ($t < 1$) suggest no effect of congruency for this measure.

First saccade direction was the second measure of search initiation to be analysed. A saccade launched in the direction of the target was defined as any saccade for which the angular direction of the saccade was within 22.5 degrees of the angular centre of a bounding box placed around the target object (as in Spotorno et al., 2014). The proportion of first saccades directed toward the target showed much greater variation across gaze cue conditions. When compared to the person absent condition, both the incongruent, $\beta = 0.050$, $SE = 0.021$, $t = 2.306$, and the congruent gaze cue conditions, $\beta = 0.062$, $SE = 0.022$, $t = 2.798$, showed a significantly higher proportion of first saccades directed toward the target. This is shown below in Figure 18.

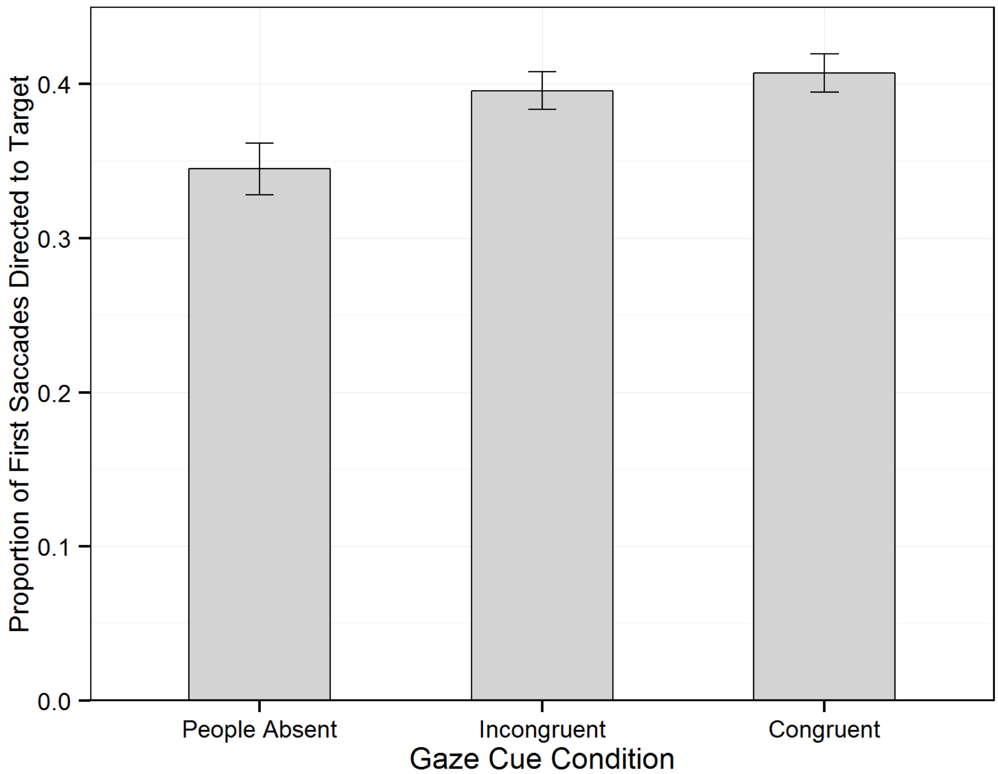

Figure 18. The proportion of first saccades directed toward the target across three gaze cue conditions. Error bars show standard error across all data samples.

Follow up analysis comparing the person present gaze cue conditions to each other found no significant differences in the proportion of first saccades directed toward the target ($t < 1$).

The final measure of search initiation was the end point accuracy of the first saccade, which describes the distance from the landing point of the saccade to the centre of the target. There were clear effects of person presence on accuracy, as can be seen in Figure 19. The end point accuracy of the first saccade was significantly improved in both the incongruent, β = -0.829, *SE* = 0.231, *t* = -3.584, and congruent, β = -1.131, *SE* = 0.228, *t* = -5.755, gaze cue conditions compared to the person absent condition. When the person present gaze cue conditions were compared, it was found that a congruent gaze cue resulted in a more accurate first saccade, β = 0.486, *SE* = 0.183, *t* = 2.656, than the incongruent gaze cue condition.

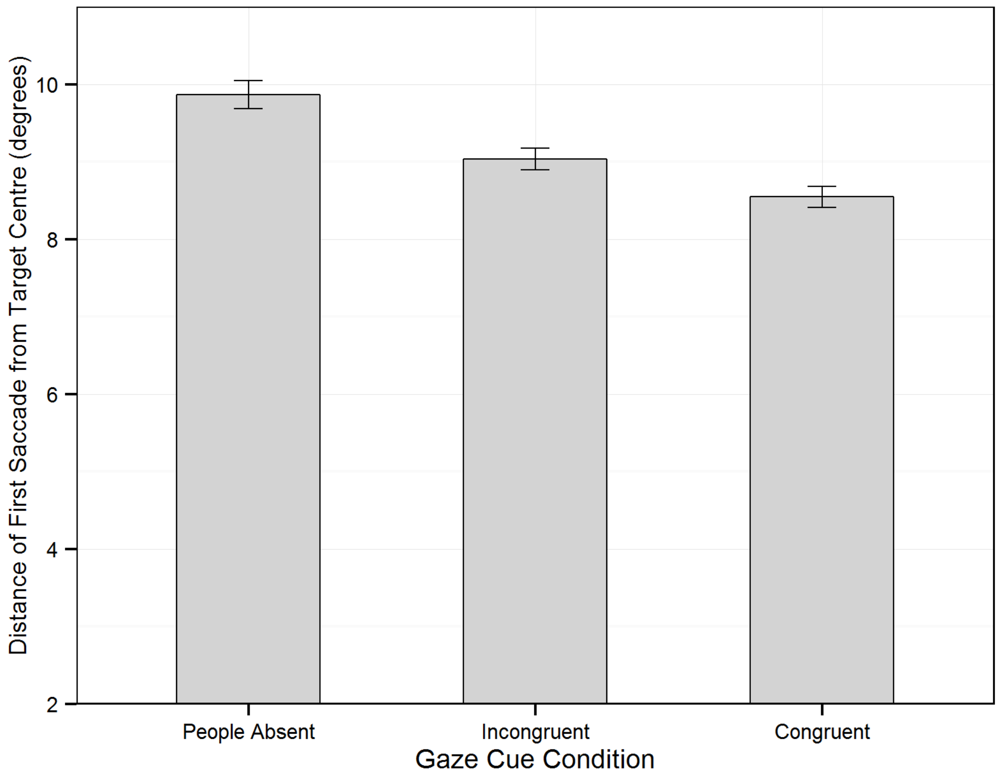

Figure 19. The distance of the landing point of the first saccade from the centre of the target ROI (in degrees of visual angle) as a measure of end point accuracy across three gaze cue conditions. Error bars show standard error across all data samples.

Scene Scanning

After initiation of search several measures were employed as a means of understanding scene scanning behaviour. First, the time taken to fixate on the target was considered, which provides a measure of how quickly the observer successfully locates the target even if they did not press the trigger button at that time. However, manual response times were also measured, which provide a broad measure demonstrating how long participants took to locate the target and decide it was the correct object, indicating search completion by pressing a trigger button on the gamepad. Following measures of speed, scan path ratio provides a measure of overall search efficiency by taking the length of the actual search route (the sum of all saccade amplitudes) and dividing it by the optimal route (the distance from the starting point of search to the centre of the target).

Error rate was also used to determine the success rate of participants. In this study, an error was defined as when the participant made a false-positive response; that is, pressing the trigger button on the gamepad when no fixation on the target object had occurred. As a measure of error it is possible that this misses some aspects of search and detection, namely that people can find and identify targets without fixating them (Henderson et al., 1997). The design of the study limits the extent to which this peripheral identification can be accounted for. Fixations that landed close to, but not within, the target boundary box may have allowed participants to identify the target, as is clear from Henderson et al.'s (1997) study that identification was very accurate without foveal vision. However, with eye tracking information being the only data gathered from search, it is impossible to tell whether object recognition occurred within these parafoveal fixations. Therefore, the strict criteria of search success is required – we can only be sure the participant has correctly located the target when they have both fixated it and pressed the gamepad button to indicate search has been concluded.

As a final point, the extent to which there was any overt orienting to, and selection of, gaze information within the search was investigated by analysing the number of fixations on the person's face within the scene. This was again determined as a fixation which landed within the boundary box drawn around the person's head. We did not include any fixations that landed on the body within this measure, as these fixations could not clearly be defined as gaze-seeking behaviour.

Person presence in the scene was of great benefit to the time taken to first fixate on the target, as can be seen in Figure 20. To satisfy model assumptions logarithmic transformation of the data was required. The LMM showed dramatic reductions in time to first fixation in both the incongruent, $\beta = -0.086$, $SE = 0.011$, $t = -7.22$, and congruent, $\beta = -0.115$, $SE = 0.011$, $t = -9.64$, gaze cue conditions.

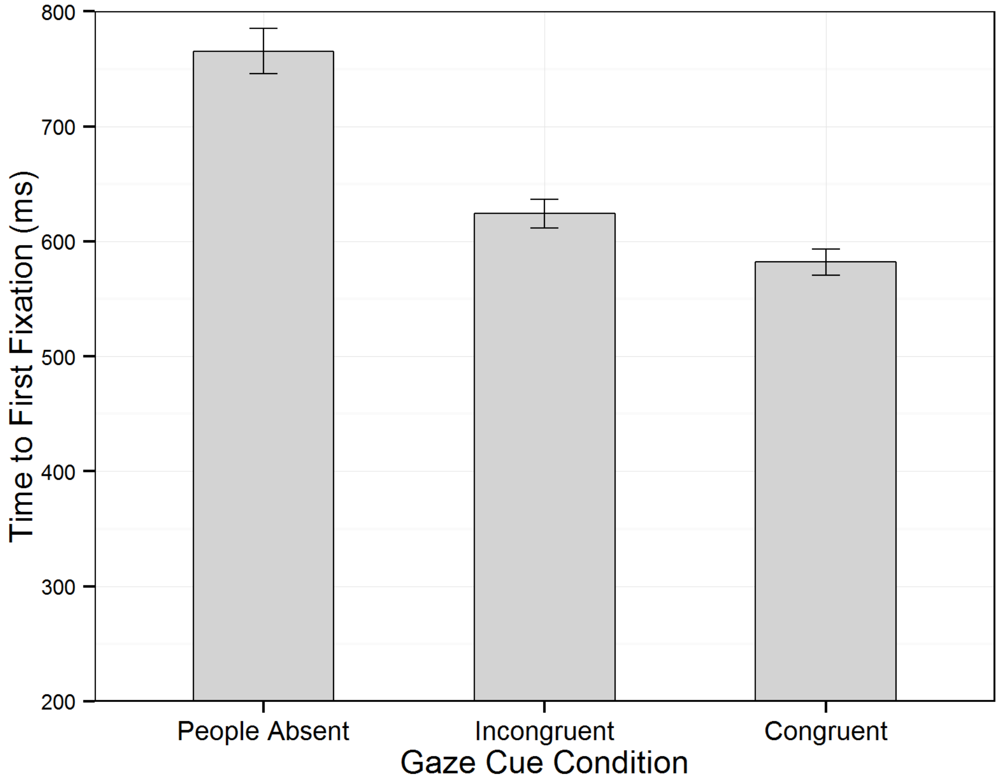

Figure 20. The time to first fixation on the target (ms) from scene presentation across three gaze cue conditions. Error bars show standard error across all data samples.

Follow up analysis showed there were considerable differences in time to first fixation on the target between person present conditions as well. When compared, the congruent condition produced significantly faster first fixations than the incongruent condition, $\beta = 0.028$, $SE = 0.009$, $t = 3.03$.

Search speed was further explored through the response time measure. As with the time to first fixate measure, response times required a logarithmic function to transform the data into a normal distribution. Figure 21 displays the considerable improvements in search speed in the person present conditions.

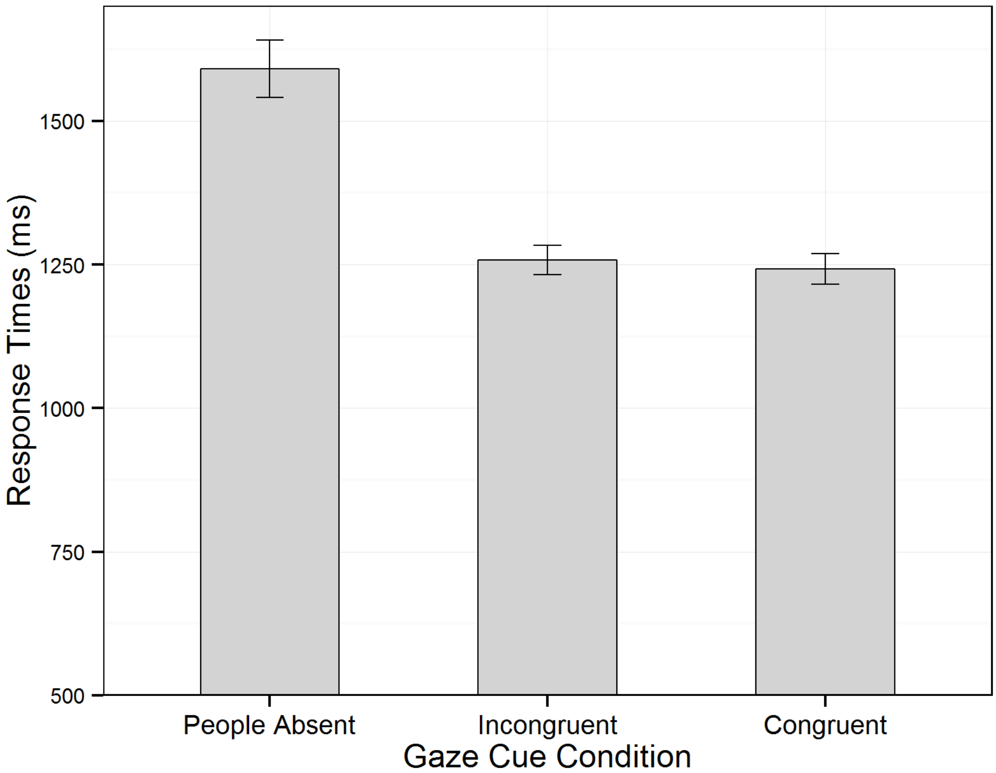

Figure 21. Response times (ms) to button press indicating successful search for the target across three gaze cue conditions. Error bars show standard error across all data samples.

Both the incongruent, β = -0.080, *SE* = 0.008, *t* = -9.17, and the congruent gaze cue conditions, β = -0.090, SE = 0.008, *t* = -10.33, showed considerably reduced response times in comparison to the person absent condition. While there was some further improvement in response times in the congruent condition as compared to the incongruent condition, β = 0.010, *SE* = 0.007, *t* = 1.42, this did not reach significance.

Considering search efficiency, an optimal route would be a single saccade from the fixation point at scene onset to the centre of the target object; therefore more efficient searches should have a scan path ratio closer to 1. Although in this model the data were not normally distributed no transformation was performed as for search efficiency it is to be expected that the majority of responses fall at the lower end of the scale with a lower ratio. Analysis showed there was improvement in search efficiency when a person was present in the scene for both incongruent, β = -0.125, *SE* = 0.067, *t* = -1.859, and congruent gaze cue conditions, β = -0.271, *SE* = 0.070, *t* = -3.847, though only the latter reached significance. This showed a significant improvement in search efficiency in the congruent gaze cue condition as compared to the incongruent condition, β = 0.145, *SE* = 0.055, *t* = 2.636. These results are presented below in Figure 22.

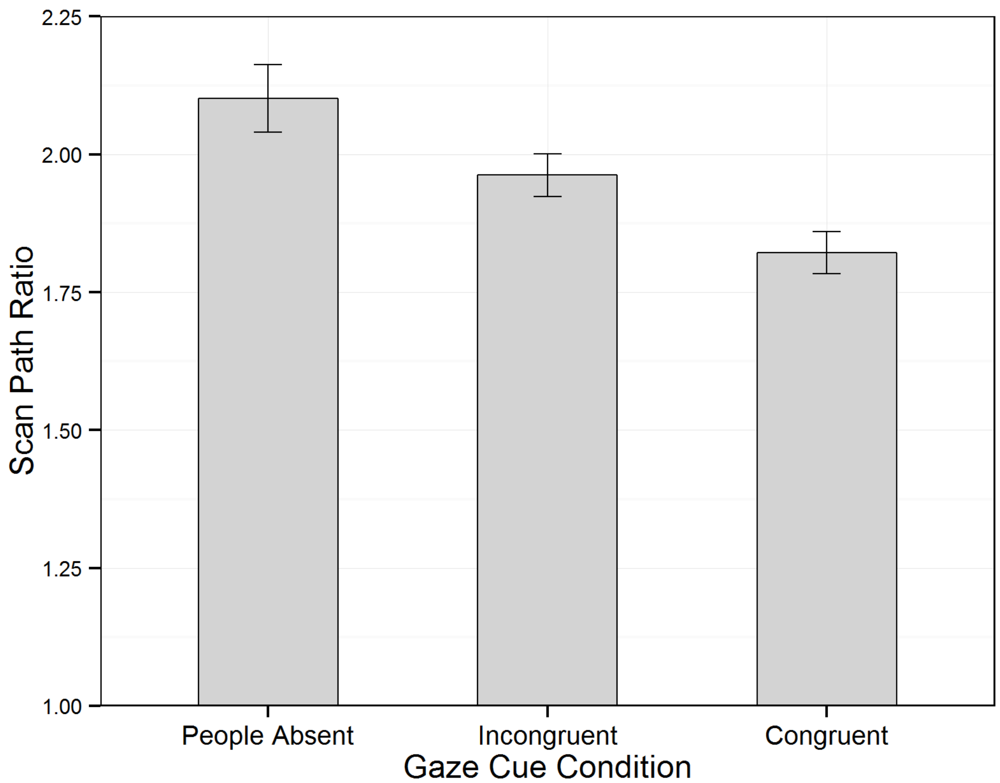

Figure 22. The scan path ratio across three gaze cue conditions. Error bars show the standard error across all data samples.

With a definition of error as the proportion of trials in which participants made false-positive responses, the error rate is already very low in baseline person absent trials. As might be predicted from Figure 23, the proportion of responses in which an error was made across all gaze cue conditions is very low, and there are no differences between either the person present conditions and the person absent condition, which required removal of correlation parameters for analysis (*t*s < 1), nor between the congruent and incongruent gaze cue conditions (*t* < 1).

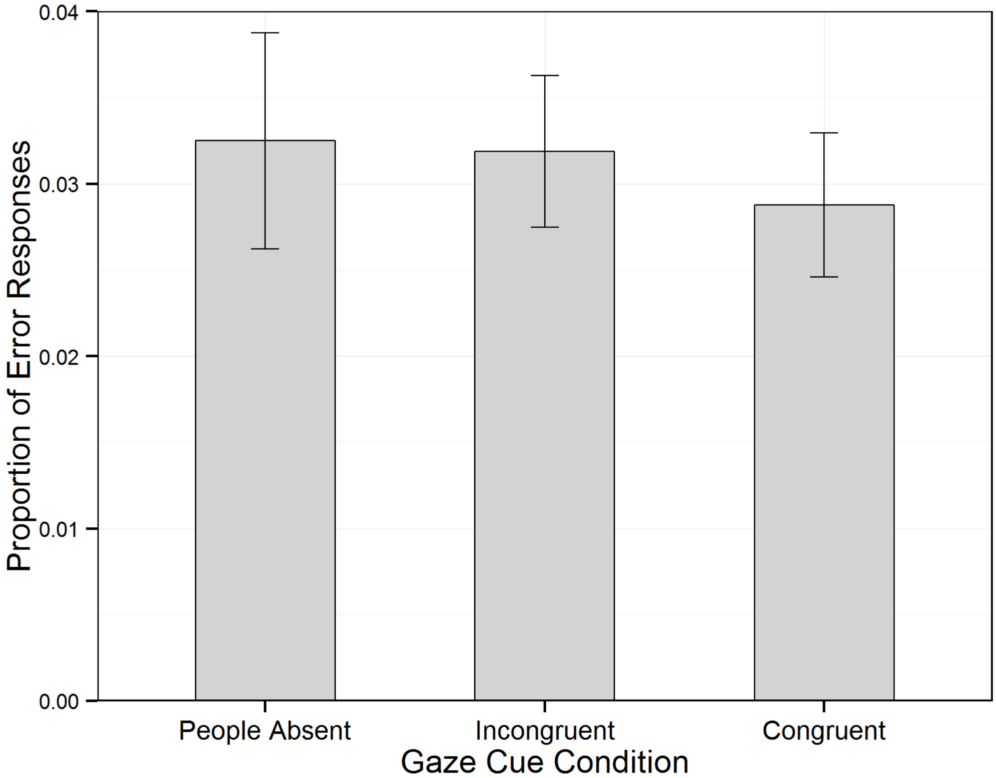

Figure 23. The proportion of false-positive responses in all trials across three gaze cue conditions. Error bars show the standard error across all data samples.

Finally, to examine the extent to which overt fixations on the person in the scene occurred, an analysis was run on ‘correct’ trials; trials where no false-positive response occurred. This is simply because in trials where participants make an error response, it is impossible to be sure of what type of strategy they were using for the search or what might have caused them to make a false-positive response. To measure the extent of overt gaze seeking, the total number of looks towards the person in the scene was calculated, with a ‘look’ defined as a fixation on the head of the person. Analysis showed that across a total of 3102 correct-response trials where a person was present, only 126 involved a fixation on the person’s face, which accounts for just 4.06% of trials. This would suggest that the results described above are a product of covert, rather than overt, gaze following.

Helpful Instruction

Search Initiation

As in the previous results section, the first saccade latency analysis for the helpful instruction condition required two transformations. Very short latencies – defined as those less than 100ms and most likely the result of pre-emptive eye movements launched prior to scene presentation – were removed (537 trials, 13.6% of the total number of trials). These data are shown in Figure 24, plotted prior to undergoing logarithmic transformation for analyses. When comparing the person present conditions to the person absent gaze cue condition, correlation parameters were removed in order to achieve convergence. Analysis showed very little difference between the three gaze cue conditions with neither the incongruent, β = 0.020, SE = 0.014, t = 1.39, nor the congruent gaze cue condition, β = 0.004, SE = 0.014, t = 0.34, showing any difference in first saccade latency when compared to the person absent condition. Similarly, further analysis showed no difference in first saccade latency between the congruent and incongruent gaze cue conditions ($t < 1$).

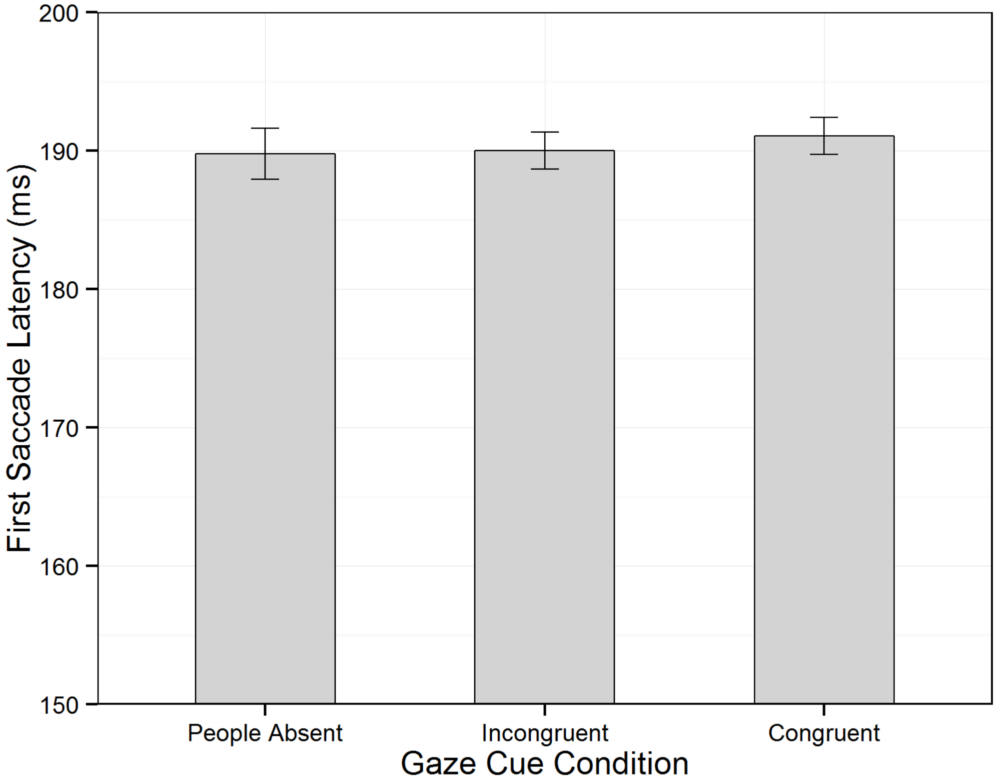

Figure 24. First saccade latency in each of the three gaze cue conditions. Error bars show standard error across all data samples.

The first saccade direction data showed greater variance across gaze cue conditions. Both the incongruent, $\beta = 0.050$, $SE = 0.021$, $t = 2.306$, and congruent gaze cue conditions, $\beta = 0.062$, $SE = 0.022$, $t = 2.798$, produced significantly higher proportions of first saccades directed toward the target than the person absent condition. Further analysis comparing the person present conditions to each other showed no difference in the proportion of first saccades directed toward the target between the congruent and incongruent gaze cue conditions ($t < 1$). These results are shown in Figure 25.

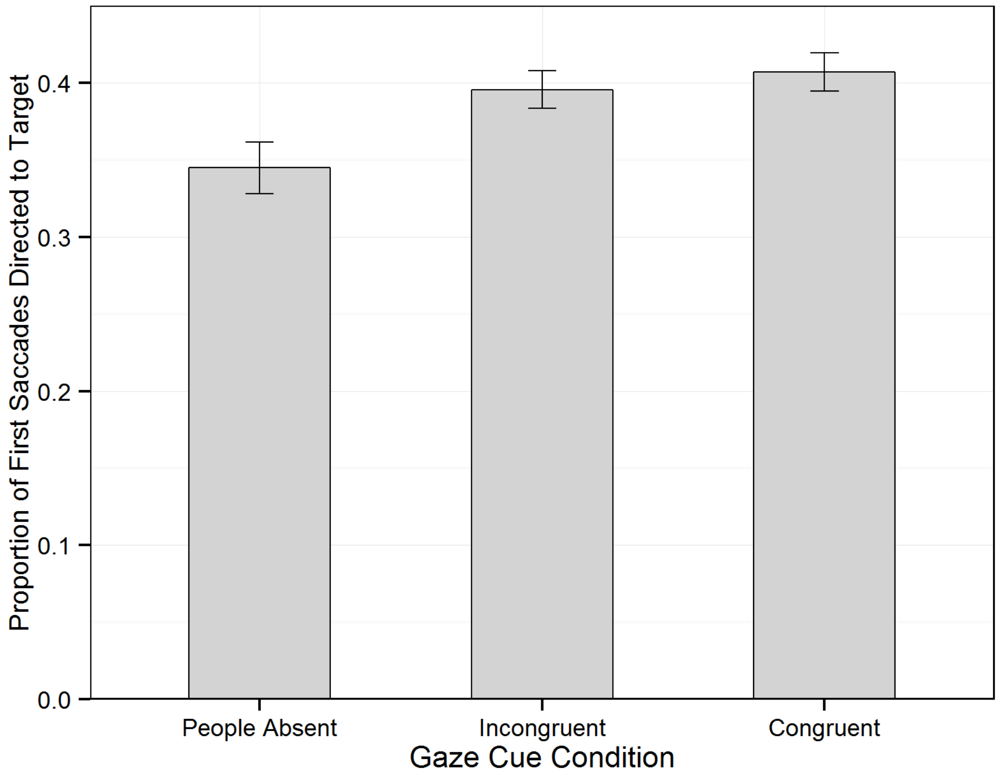

Figure 25. The proportion of first saccades directed toward the target across three gaze cue conditions. Error bars show standard error across all data samples.

The end point accuracy of the first saccade measure shows considerable differences between the three gaze cue conditions. Both the incongruent, β = -0.829, *SE* = 0.231, *t* = -3.584, and the congruent gaze cue conditions, β = -1.315, *SE* = 0.228, *t* = -5.755, produce significantly more accurate first saccades than the person absent condition. Comparing the person present conditions to each other shows further improvement when the gaze cue is congruent, β = 0.486, *SE* = 0.183, *t* = 2.656, compared to an incongruent cue. These data are displayed in Figure 26.

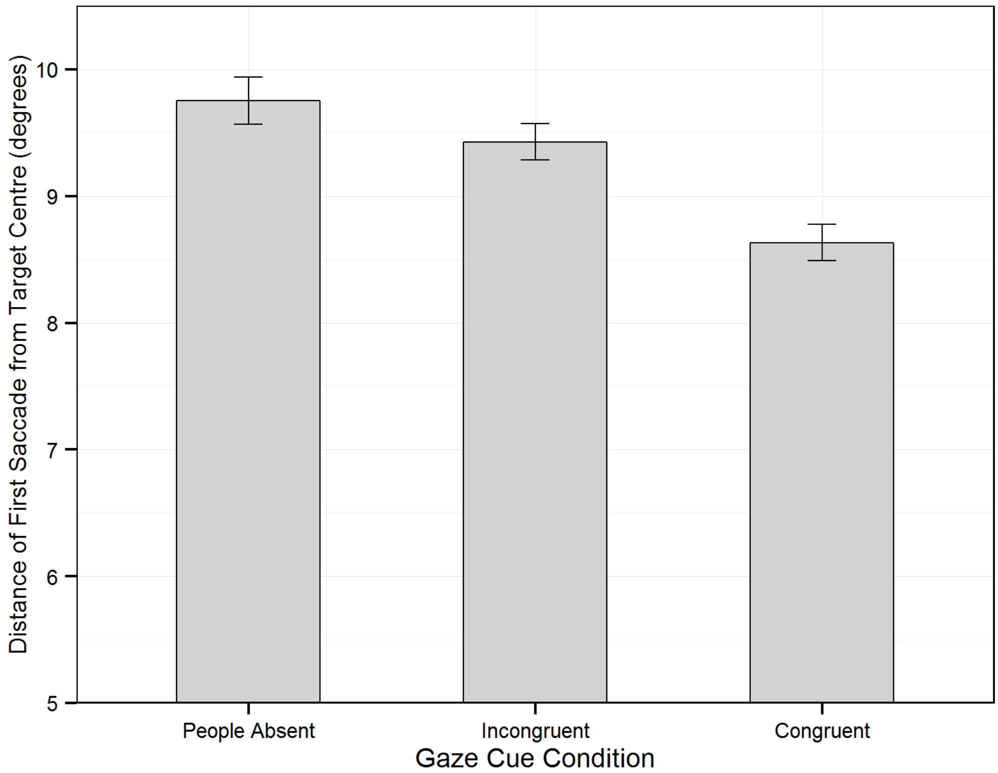

Figure 26. The distance of the landing point of the first saccade from the centre of the target ROI (in degrees of visual angle) as a measure of end point accuracy across three gaze cue conditions. Error bars show standard error across all data samples.

Scene Scanning

The first measure of scene scanning is the time taken to first fixate on the target object. To satisfy model assumptions logarithmic transformation of the data was required. There was a very strong effect of person presence on first fixation time. In both the incongruent, β = -0.086, SE = 0.011, t = -7.22, and congruent gaze cue conditions, β = -0.115, SE = 0.011, t = -9.64, time to first fixation on the target was significantly faster than in the person absent condition. This is shown in Figure 27. Further analysis comparing the person present gaze cue conditions showed further differences. When given a congruent cue participants fixated on the target for the first time much faster than when given an incongruent cue, $\beta = 0.028$, $SE = 0.009$, $t = 3.03$.

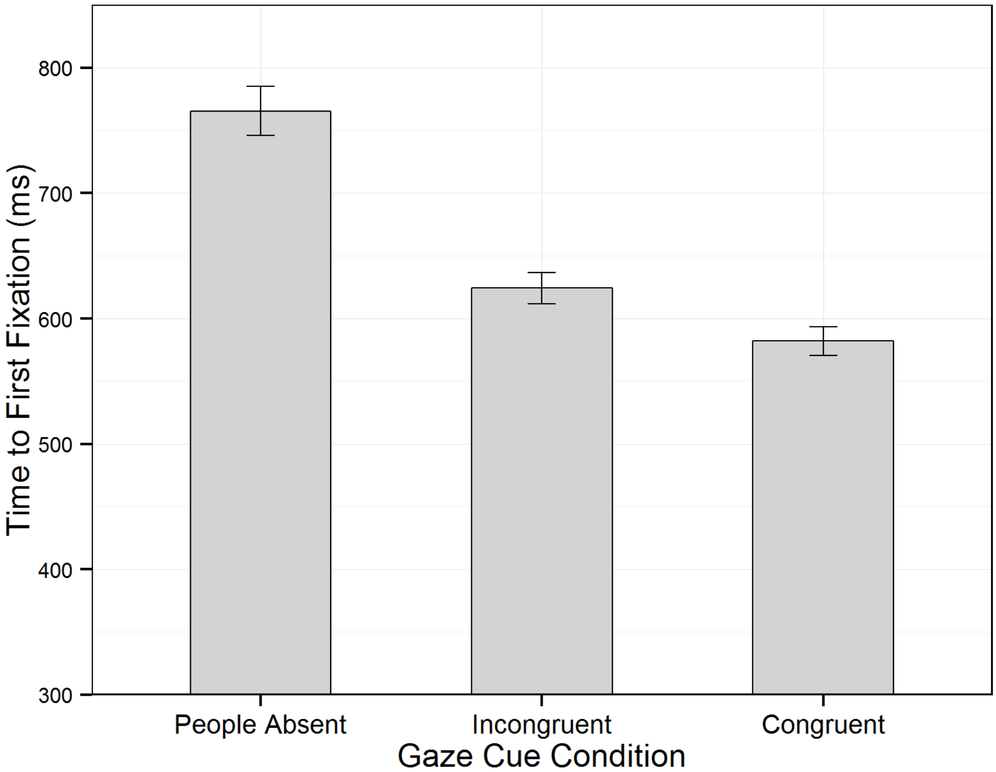

Figure 27. The time to first fixation on the target (ms) from scene presentation across three gaze cue conditions. Error bars show standard error across all data samples.

Response times provided an overall measure of how long it took participants to find the target, indicated by the point at which they pressed the trigger button on the gamepad. As with the time to first fixate measure, response times required a logarithmic function to transform the data into a normal distribution. Analysis showed considerable reductions in response time when a person was present in the scene, which can be seen in Figure 28 below. Both the incongruent, $\beta = -0.080$, $SE = 0.008$, $t = -9.17$, and congruent gaze cue conditions, $\beta = -0.090$, $SE = 0.008$, $t = -10.33$, produced significantly faster response times than the person absent condition. Further analysis showed that while there was some improvement in response times in the congruent condition as compared to the incongruent condition, $\beta = 0.010$, *SE* = 0.007, $t = 1.42$, this did not reach significance.

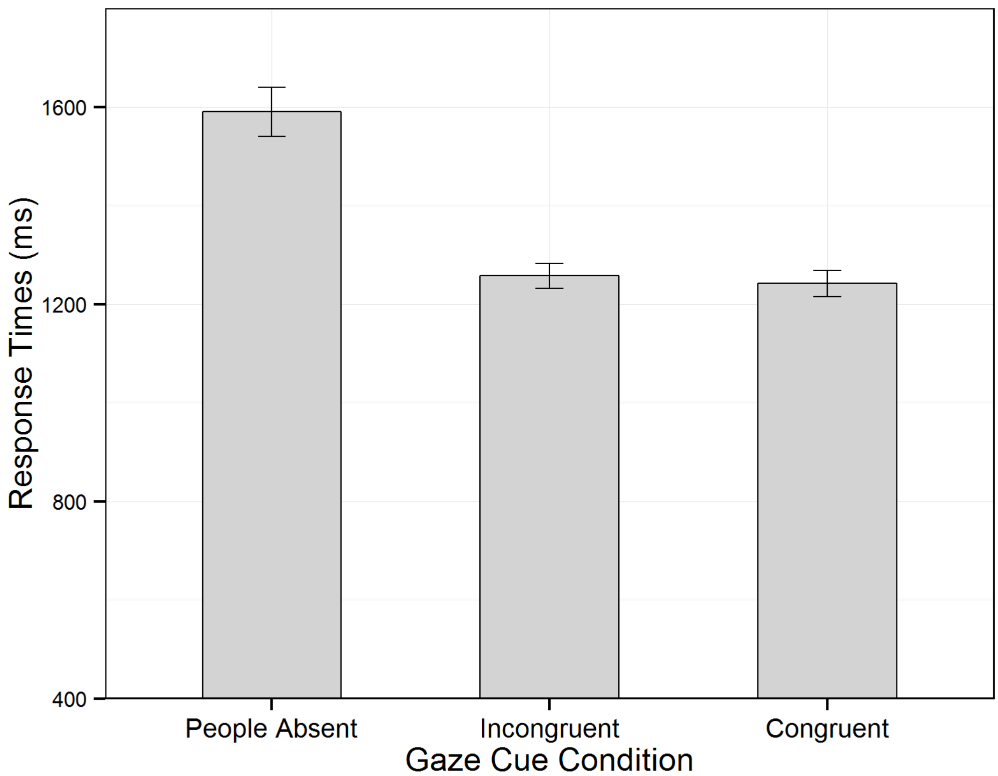

Figure 28. Response times (ms) to button press indicating successful search for the target across three gaze cue conditions. Error bars show standard error across all data samples.

When examining scan path ratio the effects of person presence seem to be mixed, as can be seen in Figure 29. Compared to the person absent condition, the incongruent condition was numerically but not significantly more efficient, $\beta = -0.125$, *SE* = 0.067, $t = -1.859$, whereas the congruent condition shows a clear benefit to search efficiency as a result of person presence, $\beta = -0.271$, *SE* = 0.070, $t = -3.848$.

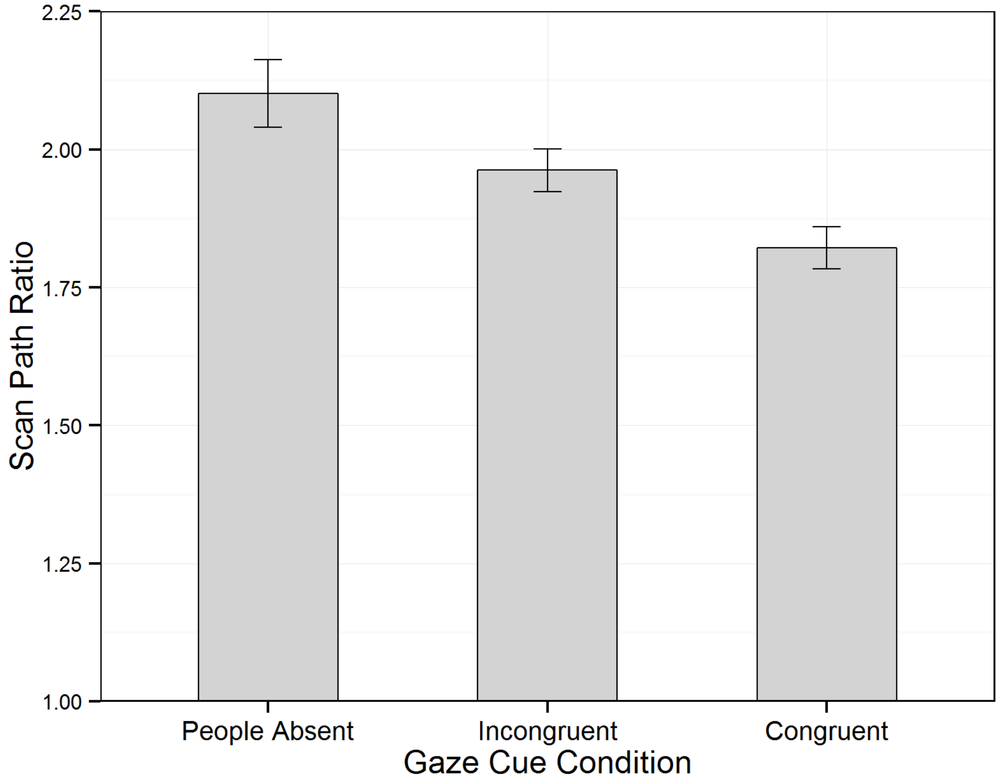

Figure 29. The scan path ratio across three gaze cue conditions. Error bars show the standard error across all data samples.

Further analyses showed that the congruent condition was also significantly more efficient than the incongruent condition, $\beta = 0.145$, $SE = 0.055$, $t = 2.636$, with a scan path ratio much closer to one.

With a definition of error as the proportion of trials in which participants made false-positive responses, the error rate is already very low in baseline person absent trials. As might be predicted from Figure 30, the proportion of responses in which an error was made across all gaze cue conditions is very low, and there are no differences between either the person present conditions and the person absent condition, (ts < 1), nor between the congruent and incongruent gaze cue conditions (t < 1).

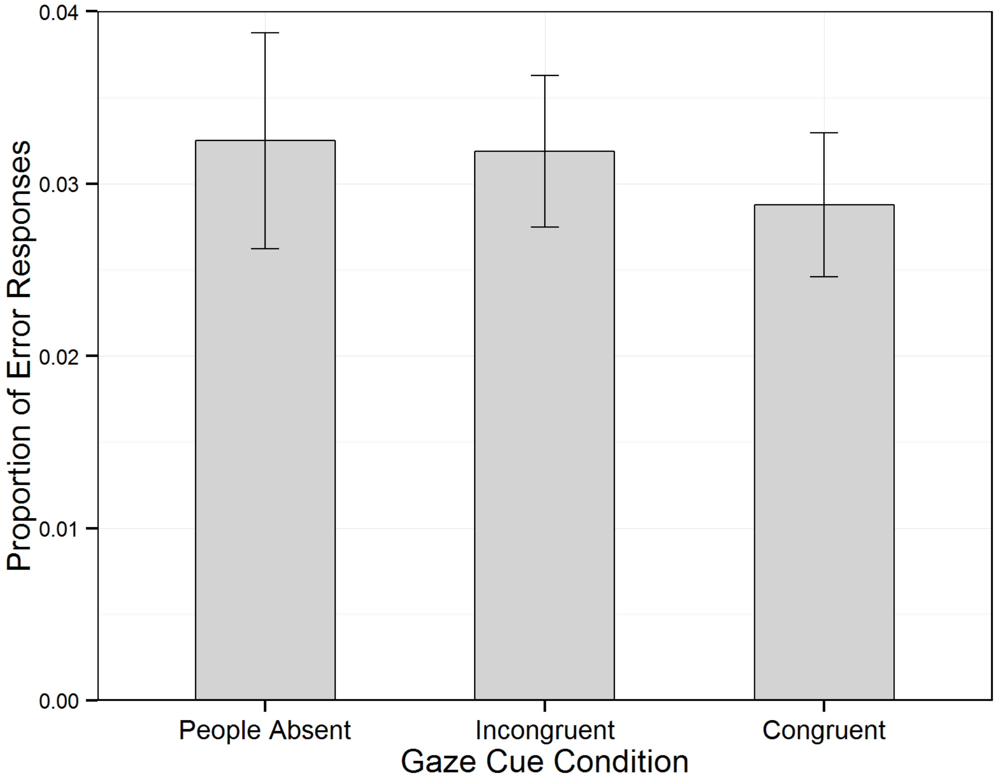

Figure 30. The proportion of false-positive responses in all trials across three gaze cue conditions. Error bars show the standard error across all data samples.

The final measure of overt gaze seeking uses the calculation of the total number of looks toward the person in the scene, only including fixations on the face as fixations on the body may not necessarily be evidence of gaze seeking. Analysis showed that across a total of 3199 correct-response trials where a person was present, only 194 involved a fixation on the person's face, which accounts for just 6.05% of trials. This would suggest that the results described above are a product of covert, rather than overt, gaze following.

Discussion

The studies documented in this chapter aim to add a new dimension to the novel search task by manipulating the instruction given to participants and examining how this impacts on their subsequent eye movements during search. The task is informed by Posner-type studies, presenting a central cue to one or other side of the scene. However, in efforts to retain more ecological validity and to more accurately represent real world gaze cues, the person present scenes show images of natural, uninstructed gaze cues that use combined head and eye movement. In the previous chapter the task was presented without giving participants any information regarding person presence, and found that while person presence was beneficial to performance in the search task across almost all measures, there were no effects of congruency found. The current chapter looks to investigate how observer search behaviour changes based on the additional instruction they are given regarding person presence. In the unhelpful instruction condition participants were told that the person was not relevant to the task, replicating the instruction format used in many of the previous Posner-type tasks. The helpful instruction condition informed participants that the person may be looking at the target and therefore may help them find the target faster. These manipulations aim to change the participants' perception of the usefulness of the cue provided within the scene.

Considering first the search initiation phase, the effects of person presence are mixed. First saccade latency produced the smallest variation in performance across the three gaze cue conditions for both helpful and unhelpful instruction. There were no effects of person presence or congruency on pre-launch processing in either instruction type. Effects only began to emerge after the first saccade had been launched in measures that address first saccade accuracy: the direction of the first saccade and its end point accuracy. In both instruction conditions the proportion of first saccades directed toward the target was much higher once a person was present in the scene and these saccades were more accurate, bringing the eyes closer to the centre of the target compared to person absent trials. First saccade direction was unaffected by the congruency of the gaze cue, but the end point accuracy of the first saccade was considerably improved when participants were given a congruent cue rather than an incongruent cue. These results suggest that in the search initiation phase, person presence benefits performance only after the first saccade is launched. The congruency of the gaze cue provided by the person in the scene does not have much effect on the likelihood of the saccade being directed toward the target, but does significantly improve the accuracy of the first saccade with congruent gaze cues producing the highest accuracy of all three gaze cue conditions.

In the scene scanning stage of search, the effects of person presence are clearer and identical across both instruction conditions. Speed of search – as measured by both the time taken to first fixate on the target and the overall response time – was considerably improved once a person was present in the scene. While response time showed no congruency effects, the first fixation on the target occurred faster in the congruent gaze cue condition. The effects of congruency in the measure of search efficiency – scan path ratio – are more subtle. In this measure person presence only provided search efficiency with a significant boost when a congruent gaze cue was given. Incongruent gaze cues produced no real differences in efficiency than person absent scenes. That means that for this measure, congruent gaze cues resulted in more efficient search than both incongruent gaze cues and person absent scenes.

One of the first studies to argue for reflexive orienting to gaze cues was conducted by Driver et al. (1999) who performed three experiments to determine whether gaze cues would result in facilitated performance in a letter discrimination task, even when these cues had no bearing on what letter would appear. The centrally presented face looked either left or right, and a T or L then appeared to the left or right of the face. Participants were asked to discriminate which letter had appeared as quickly as possible. While the gaze cue did not offer any information regarding the letter identity, Driver et al. (1999) hypothesised that if reflexive orienting did indeed occur, discrimination would occur faster for letters that had appeared in the area cued by the face. This was found to be the case in all three experiments. Crucially, Driver et al. (1999) controlled for several factors which could have otherwise explained this result. In experiments one and two the gaze cue was equally likely to be given to the left or the right, and was therefore spatially uninformative – following the cue offered no greater benefit than ignoring it. This means that effects of orienting found in these experiments are not due to any spatial cuing effect. Furthermore, in experiment three Driver et al. (1999) manipulated the gaze cues to be incongruent with target location four times more frequently than they would be congruent with target location. Participants were advised of this at the beginning of each experimental trial. The fact that the cuing effect was not disrupted by this change in the probability of a congruent cue demonstrates quite conclusively that cuing facilitates performance at congruent locations, even when participants have been advised the cues will be unhelpful.

As in Driver et al.'s (1999) first two experiments, the studies reported in this chapter were spatially uninformative with gaze cues equally likely to occur in a direction that was congruent with target location as it would be to provide an incongruent cue. Although participants' eye movements have been explored in more detail in these studies with a wider range of measures, the findings seem to some extent to support a hypothesis of reflexive orienting. The improved performance seen when a person is present fits with the hypothesis that participants will follow the gaze of the person in the scene even when this is spatially uninformative and they are told it is not useful.

The results of the current studies also conform to findings from Ricciardelli et al. (2002) who again used Posner-type tasks to determine whether reflexive orienting in response to gaze cues occurred. In the first study, Ricciardelli et al. (2002) were testing their hypothesis that social stimuli – in this case, gaze cues – elicited unique responses in observer eye movements. Testing both static and dynamic gaze cues, the authors found that observers display a tendency to mimic the gaze direction of others, thus engaging in joint attention, but not to follow the cue of non-gaze cue stimuli (e.g. arrows). In particular relevance to the current studies, Ricciardelli et al. (2002) made it explicitly clear to participants that the gaze cues were not helpful to the task, aligning it most closely to the unhelpful instruction condition discussed above.

The current study shows no effect at all of instruction: participants' performance and eye movement behaviours are identical across the helpful and unhelpful instruction conditions. Does this mean then that regardless of what participants are told the automatic effects of gaze cues occur? Recent research suggests this may be the case. Greene, Mooshagian, Kaplan, Zaidel and Iocaboni (2009) conducted a spatial cuing task while monitoring participants' brain activity with fMRI. They reasoned that while the Posner-type tasks discussed previously demonstrate reflexive orienting to social cues, their comparison to non-social cues have only used symbolic cues that engage top-down processes. To properly determine whether gaze elicited a genuinely unique response in observers, Greene et al. (2009) used the fMRI to monitor neural activity when social or non-social cues were presented over long and short stimulus onset asynchronies (SOAs) in a Posner-type task. Both social and non-social cues facilitated participant performance at the short SOA, and the social cues also had an inhibitory effect at this SOA when the cue was incongruent. However, inhibition of return (IOR) was only found for non-social cues when participants were presented the stimulus at a long SOA. There were also differences in the brain regions activated in response to the different types of cue. Social cues resulted in much greater activation in the occipito-temporal regions regardless of the SOA used, while non-social cuing only demonstrated greater subcortical activity when at the long SOA. As there was some overlapping in behavioural responses and neural activation across social and non-social cues Greene et al. (2009) conducted a control experiment to rule out spatial location of the cue as a cause for differences that did occur between the two types of cue. They followed the same methodology of a centrally-presented face, though in this control experiment directional arrows would appear in the location of the mouth of the face. These arrows were presented for the same duration as cues in the earlier experiment and similarly were not predictive of target location. This control experiment produced the same facilitation effects as the main experiments conducted by Greene et al. (2009), leading the authors to conclude that this clearly demonstrates the differences in behavioural and neural responses are due to the social nature of the gaze cue, not just its spatial location.

Greene et al.'s (2009) study provides strong neurological evidence that social gaze cues have neural regions dedicated to their processing. The authors argue that this demonstrates an ‘evolutionary trajectory’ for reflexive orienting, due to the cortical mechanisms dedicated to gaze cue processing. If the position of gaze as a special stimulus is so hard-wired in the brain, it is quite possible that task instruction cannot override this automatic process. This would explain why the current studies demonstrate effects of person presence, which can be presumed as early evidence for reflexive orienting, regardless of the instruction given to participants. Furthermore, this would contribute strong evidence for the argument of reflexive orienting, showing that these effects persist irrespective of task demands.

However, the current study does not fully support Ricciardelli et al.’s (2002) or Greene et el.’s (2009) findings. While there are clear benefits of person presence found across almost all measures, there are very few effects of congruency. If observers were engaging in joint attention with the person in the scene providing the gaze cue, there should be a deterioration of performance when an incongruent cue is provided in comparison to behaviour when a congruent cue is given. In literature that supports a reflexive orienting hypothesis the facilitation effect persists even when participants are explicitly told the gaze cues are unhelpful, which can be seen in the results of the studies described in this chapter. Were reflexive orienting occurring, there should be improvements in performance evident in the congruent condition as compared to the person absent and/or incongruent conditions, which is not wholly apparent in the results described above. However, there are some limited effects of congruency apparent in the results of these studies, seen in measures of the end point accuracy of the first saccade, the time to first fixation on the target, and of search efficiency. It should be remembered that Posner-type studies have not examined participant eye movement behaviour with as wide a range of measures as have been used in the current studies. A lack of congruency effects in some (but not all) measures is not necessarily reason to discount these studies as support for a reflexive orienting hypothesis: that conclusion cannot be drawn without evidence of congruency effects – or lack thereof – in the corresponding measure in previous studies. It is possible that had these measures been investigated in previous Posner-type literature, the effects of congruency may not have been as clear-cut as they appear to be in the measures that are presented. For the sake of accuracy, it cannot be said that the results of the current studies fully support a reflexive orienting hypothesis because the congruency effects that would be predicted by this model are not apparent in all measures, however, that does not discount that in some measures congruency effects are beginning to emerge and the overall trend of results – which demonstrate strong benefits of person presence – do fit with a reflexive orienting model.

The results are also somewhat consistent with a social facilitation model, as discussed in the previous chapter. Richardson et al. (2012) would argue that perceiving something alone is inherently different from perceiving it as the same time as another person. To investigate how pervasive this effect is, and whether it would persist even with the most minimal context of social interaction, the authors presented images to participants and manipulated whether they believed another unseen person was also looking at the same sets of images. Their participants showed a clear preference for images they believed had also been looked at by another unseen person: these ‘shared’ images were remembered better and looked at differently – there were no increases in overall looking time compared to images looked at alone, but the distribution of fixations across the scene changed when participants believed another person was also looking at the image. Considering the stimuli presented to participants in the current studies, they would clearly be able to see when the person in the scene was attending the same object as them. One might expect that social facilitation should only occur in trials where the person is providing a congruent gaze cue, thus engaging in joint attention. . It is possible that there is a broad social facilitation provided by the mere presence of an individual (as in Allport, 1920; Zajonc, 1965), but the more specific benefits of joint attention, as discussed by Richardson et al. (2012) are only beginning to emerge in some measures. While the general consensus of social facilitation is that the presence of another person can improve an individual's performance in a task, there are studies that demonstrate the presence of another person can hinder performance too. Markus (1977) asked participants a simple task for which there was no performance criteria so as to avoid evaluation apprehension – participants simply had to dress and undress in familiar and unfamiliar clothing. They completed this task alone, in the presence of a person who was inattentive, or in the presence of an active spectator. Markus (1977) found that both observed conditions (i.e. with a passive or active spectator) improved performance when dressing with familiar clothing, but hindered performance when dressing with unfamiliar clothing. In terms of the current studies, Markus' (1977) findings demonstrate that there may be trials in which participants are facilitated by person presence, for example if they are looking with an item they are familiar with. However, if looking for an object they are less familiar with, or a familiar object that presents in an unusual way (a mug in an unusual shape, for example) the very same stimulus of a person in the scene may cause their performance to deteriorate. This may explain why apparent social facilitation in person present scenes does not result in further performance enhancement in congruent gaze cue trials where participants would share in joint attention on the target object.

This chapter has documented observer behaviour in realistic Posner-type tasks where the instructions given to participants regarding the presence of a person in the scene were manipulated to either reflect an unhelpful instruction (the person is irrelevant), or a helpful instruction (the person may be looking at the target). Results indicate there are strong effects of person presence across all measures of search, and that there is some evidence of congruency effects occurring after the first saccade has been launched. To fully support a reflexive orienting hypothesis, stronger congruency effects would have been expected. However, these studies use a broader range of measures than were applied in previous Posner-type tasks, so the lack of congruency effects across all measures is not reason to discount these studies as at least somewhat supporting a reflexive orienting hypothesis. Considering the effects of social facilitation, evidence from Markus (1977) suggests that participants may experience both facilitatory and detrimental effects of person presence while conducting a task, depending on whether this task is familiar to them or not. It is possible that participants benefited from person presence in trials where they searched for familiar objects, but were negatively affected by person presence when searching for unfamiliar objects. Neural evidence presented by Greene et al. (2009) suggests that the special status of gaze cues in the brain is irrefutable and therefore it is possible that task instruction simply cannot override the cognitive processes dedicated to processing this type of stimuli, which may explain why no difference in participant performance was observed between the helpful and unhelpful instruction conditions. However, it is also possible that the lack of effect of instruction may indicate that the manipulation was not successful, either because participants did not believe the manipulation in the instruction, or because simply mentioning the presence of a person in the scene made them more salient regardless of the context surrounding the mention of the person. To fully understand the role of task instruction on observer eye movement behaviour, and how these instructions interact with the type of gaze cue presented, a more in-depth consideration of instruction is required. This will be the focus of the next chapter, in which I will statistically compare the three experiments discussed so far, in order to explore these issues more carefully.

Chapter Four

Comparing the effect of different instructions on observer eye movement behaviour in a single gaze cue visual search task

While previous Posner-type task literature (e.g. Driver et al., 1999; Friesen & Kingstone, 1998; Ricciardelli et al., 2002) all find strong evidence for reflexive orienting in response to gaze cues, recent literature has begun to question the ecological validity of these paradigms, and has moved toward more real world approaches (e.g. Macdonald & Tatler, 2013a, 2013b; Tatler, Kirtley, Macdonald, Mitchell & Savage, 2013; Risko, Laidlaw, Freeth, Foulsham & Kingstone, 2012). To address these concerns, whilst retaining the basic principles of providing a gaze cue that was either congruent or incongruent with target location, Chapter Two introduced a more realistic paradigm in which the cue is presented alongside the context of a real environment, a body, and a natural accompanying head direction cue. In Chapter Three, different types of instruction concerning person presence in the scenes were added to the task in order to determine how task demands influence observer eye movement behaviour (as discussed in Itier et al., 2007). While there is evidence that task instruction can have strong effects on observer eye movements (e.g. Yarbus, 1967), the results in Chapter Three seemed to suggest that observers behaved similarly regardless of the instruction they were given. However, the findings of the three studies in Chapters Two and Three were not directly compared, so conclusions about the relative influences of task demands are thus far qualitative rather than quantitative. In this chapter quantitative comparisons are made across the three studies presented so far in this thesis. To fully understand the role instruction plays, it is important to compare not only the experiments where participants were given an instruction regarding the purpose of person presence within the scene, but also to include the baseline experiment where no instruction was given.

Data Analysis

The main purpose of this chapter is to examine how instructions varied the manner in which gaze cues were responded to. To explore the influence of interactions, LMM models were run using the R statistical analysis environment (R Development Core Team, 2011). Both gaze cue condition and instruction condition were three-level factors, allowing two contrasts to be set up for each factor using the contrasts() function. For the gaze cue factor, the first contrast was set up to compare the person absent scenes to the person present scenes (combining across the two person present gaze cue conditions). The second contrast was set up to compare the congruent and incongruent gaze cue conditions to each other, ignoring the person absent condition. For the instruction type factor, the first contrast was set up to consider whether the no instruction condition differed from the instruction conditions which make person presence salient (helpful and unhelpful instruction conditions combined). The second contrast was set up to examine whether the helpful and unhelpful instruction conditions differ from each other, ignoring the no instruction condition. The LMM was set up to consider these two contrasts for each factor and their interactions (thus four interaction terms describing whether each contrast in one factor depends upon each contrast in the other factor).

The results of the contrasts and LMM interactive model are presented in a similar way to the analyses of the previous chapters with *t*-values for linear models, but it should be noted that in this chapter the statistics describe whether each coded contrast is significant rather than comparisons between individual levels of a factor as in other chapters. Again, as in previous chapters, we consider any effects for which the *t*-value is greater than two as reflecting a significant effect (as in Kleigl et al., 2012).

Results

First Saccade Latency

The first saccade latency data featured a number of very short latencies, which could only be the result of saccades programmed before the scene appeared and thus were not influenced by anything within the content of the scene. For this reason, very short latencies were removed from the dataset and the remaining latencies underwent logarithmic transformation to generate a normal distribution. Before exploring first saccade latency in each of the gaze cue conditions, the data is first presented together in Figure 31 below.

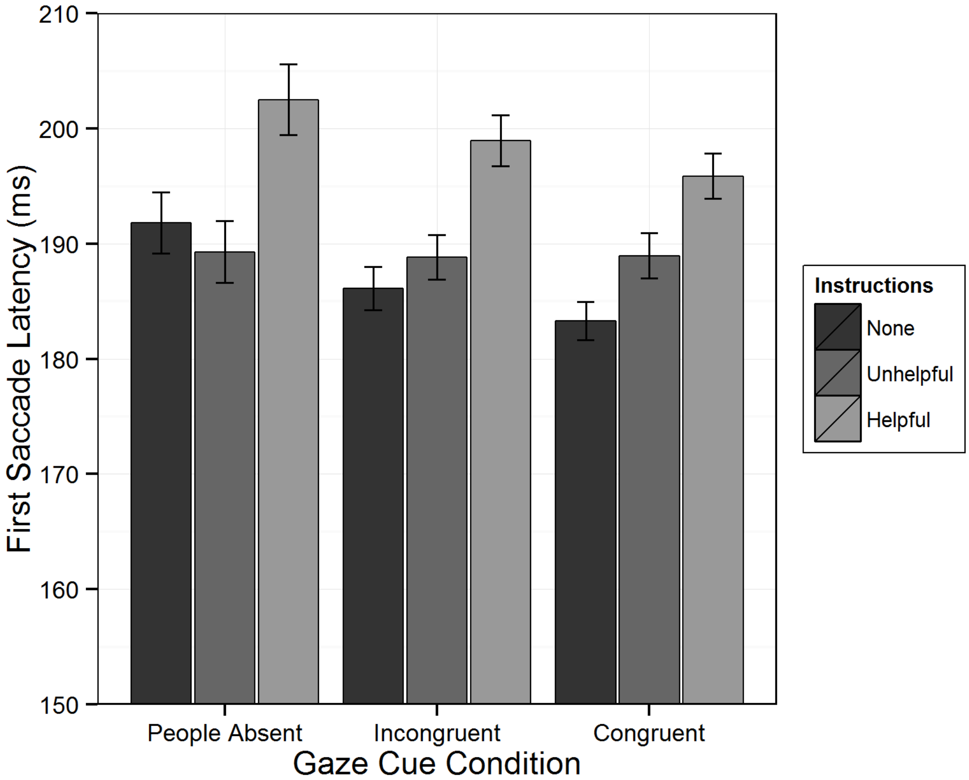

Figure 31. First saccade latency (ms) across all three gaze cue conditions, with error bars displaying standard error across all data samples. The dark grey bar represents the no instruction condition, medium grey bar represents the unhelpful instruction condition and the light grey bar represents the helpful instruction condition.

Contrasts of gaze cue conditions showed that the person absent scenes resulted in significantly longer first saccade latencies than person present scenes, $\beta = 0.009$, $SE = 0.003$, $t = 2.38$. However, there was no difference in first saccade latencies when participants were given an incongruent gaze cue as compared to a congruent gaze cue, $\beta = 0.003$, $SE = 0.003$, $t = 1.10$. When comparing instruction conditions, it was found that there were significantly longer first saccade latencies when participants were given no instruction versus when they were given some instruction regarding the presence of a person in the scene, $\beta = -0.014$, $SE = 0.003$, $t = -4.68$, (i.e. helpful/unhelpful instruction conditions). Analysis of first saccade latencies in the helpful and unhelpful instruction conditions showed that being told to ignore the person in the scene resulted in shorter latencies than when participants were told the person may be helpful, $\beta = -0.020$, $SE = 0.003$, $t = -5.76$.

Examining the interaction between gaze cue condition and instruction condition showed that first saccade latencies that occurred in person absent versus person present scenes were not dependent on whether participants were given instructions that made no reference to person presence (no instruction) or instructions that did make reference to person presence (helpful/unhelpful instruction), $\beta = 0.008$, $SE = 0.007$, $t = 1.17$. Similarly, first saccade latencies in incongruent or congruent gaze cue conditions did not vary based on instruction that did or did not refer to person presence ($t < 1$). When comparing the effects of helpful versus unhelpful instruction, it was found that instruction condition had no effect on first saccade latencies in person present gaze cue conditions compared to the person absent gaze cue condition ($t < 1$), nor on first saccade latencies in incongruent versus congruent gaze cue conditions ($t < 1$).

First Saccade Direction

For analysis of first saccade direction no transformation of the data was required. As in all previous analyses, a saccade toward the target was defined as one which falls within 22.5 degrees of the angular centre of the boundary box placed around the target. Figure 32 below shows the proportion of first saccades directed toward the target in each instruction condition across the three gaze cue conditions.

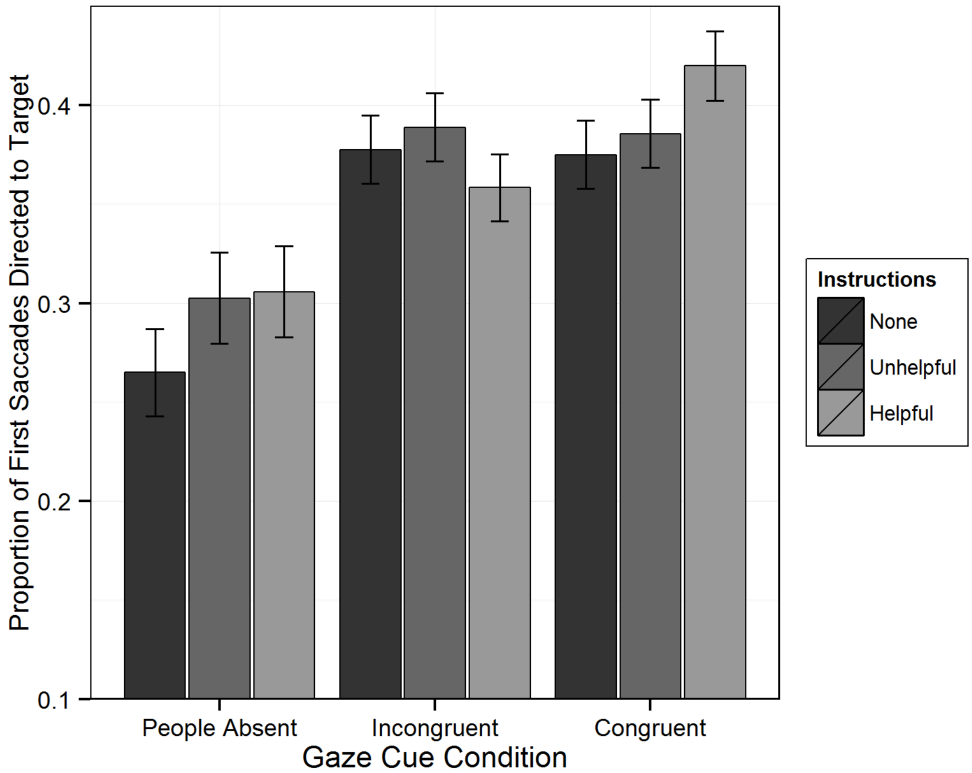

Figure 32. The proportion of first saccades directed toward the target across all three gaze cue conditions, with error bars displaying standard error across all data samples.

Analysis showed that person absent scenes resulted in significantly less first saccades made in the direction of the target than person present scenes, $\beta = -0.096$, $SE = 0.020$, $t = -4.731$, however, there was no difference in the proportion of first saccades directed toward the target in the incongruent gaze cue condition compared to the congruent gaze cue condition, $\beta = -0.018$, $SE = 0.014$, $t = -1.307$. Contrasts of instruction conditions showed that there was no difference in the proportion of first saccades directed toward the target when participants were given no instruction person presence compared to instructions that did make reference to person presence, $\beta = -0.018$, $SE = 0.013$, $t = -1.326$. There was no difference in the proportion of first saccades directed toward the target in the helpful instruction condition compared to the unhelpful instruction condition ($t < 1$).

Examination of interactions between gaze cue condition and instruction condition showed whether the instruction did or did not make reference to person presence had no effect on the proportion of first saccades directed toward the target in person absent versus person present scenes ($t < 1$), nor did it affect first saccade direction in incongruent versus congruent gaze cue conditions, $\beta = 0.031$, $SE = 0.028$, $t = 1.108$. The proportion of first saccades directed toward the target in person present versus person absent scenes was not affected by whether the instructions given did or did not make reference to person presence ($t < 0.05$). There was evidence of some difference in the proportion of first saccades directed toward the target in the incongruent and congruent gaze cue conditions depending on whether participants were given a helpful or unhelpful instruction, but this was not significant, $\beta = 0.064$, $SE = 0.032$, $t = 1.973$.

End Point Accuracy

Figure 33 displays the results of each instruction type across the three gaze cue conditions. Analysis showed there were clear effects of person presence on the end point accuracy of the first saccade, with person present scenes producing greater accuracy than person absent scenes, $\beta = 1.414$, $SE = 0.174$, $t = 8.100$. The congruent gaze cue condition produced more accurate first saccades than the incongruent gaze cue condition, $\beta = 0.373$, $SE = 0.157$, $t = 2.370$. However, there were no effects of instruction on end point accuracy of the first saccade. There was no difference in end point accuracy between instructions that did or did not make reference to person presence ($t < 1$), nor between the helpful and unhelpful instruction conditions ($t < 1$).

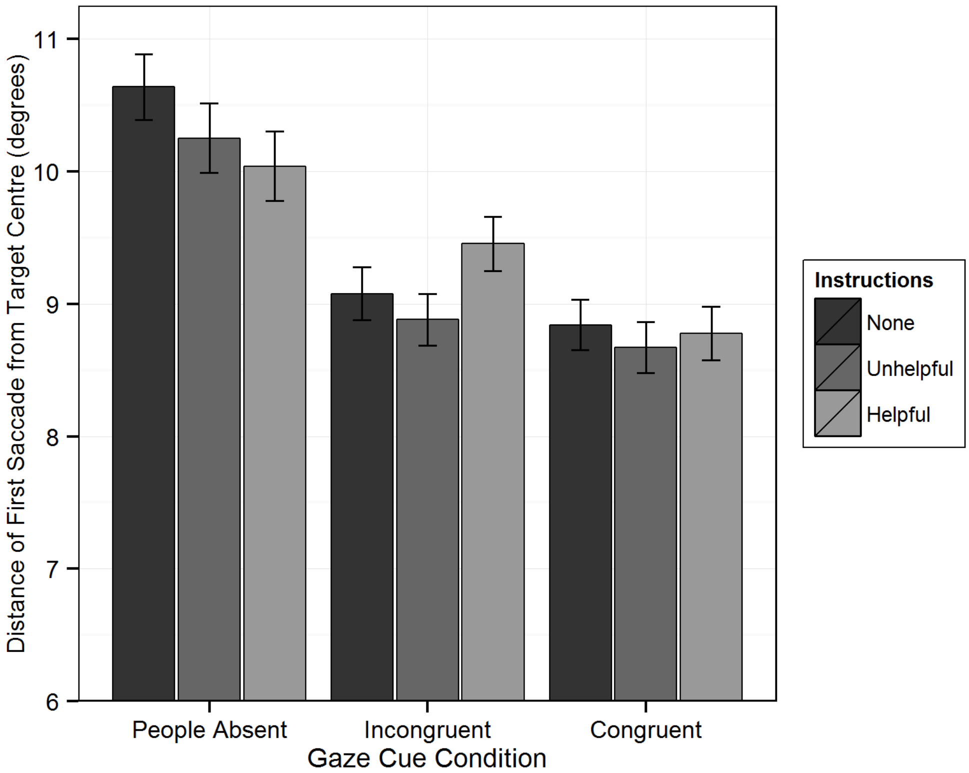

Figure 33. The distance of the landing point of the first saccade from the centre of the target boundary box (in degrees of visual angle) as a measure of end point accuracy across three gaze cue conditions. Error bars show standard error across all data samples.

Examining the interaction between gaze cue conditions and instruction conditions showed whether the instruction did or did not make reference to person presence had no effect on end point accuracy in person absent versus person present scenes ($t < 1$), nor did it affect accuracy in incongruent versus congruent gaze cue conditions ($t < 1$). Having a helpful or unhelpful instruction did not affect participants' end point accuracy in person present versus person absent scenes, β = 0.543, *SE* = 0.413, t = 1.314, nor in congruent versus incongruent gaze cue conditions, β = -0.473, *SE* = 0.365, t = -1.294.

Time to First Fixation on the Target

To satisfy model assumptions a logarithmic transformation was applied to the data, generating a normal distribution. Figure 34 below shows the data in its original form across all gaze cue conditions and instruction types.

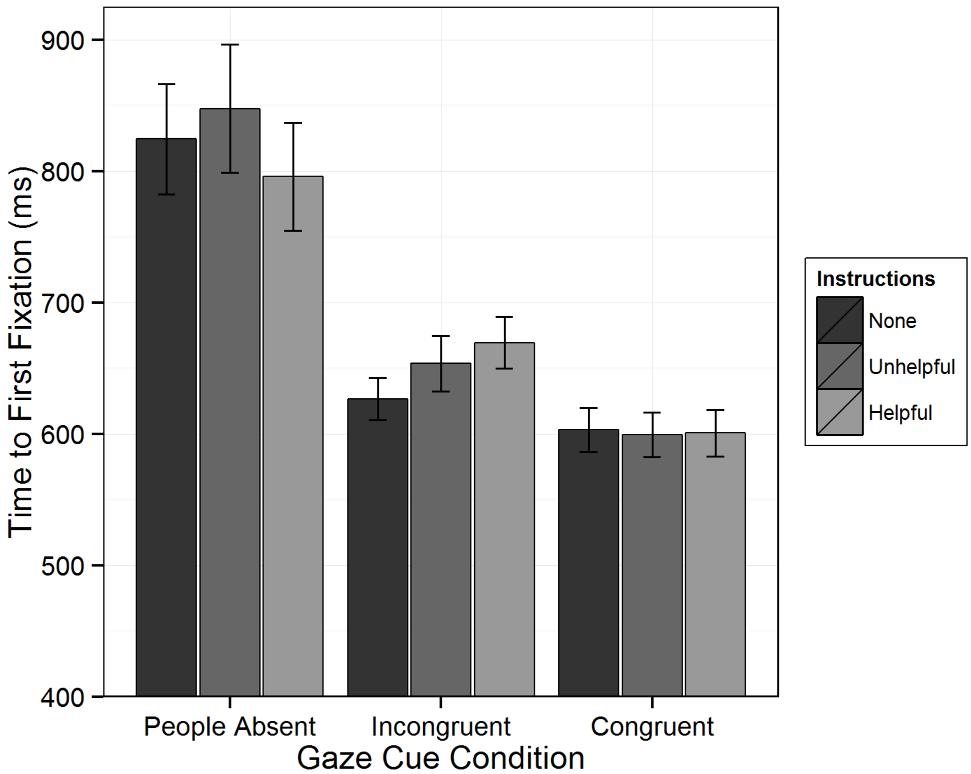

Figure 34. The time to first fixation on the target (ms) from scene presentation across three gaze cue conditions, where error bars show standard error across all data samples.

Contrasts demonstrated that the time to first fixation on the target was significantly longer in person absent scenes than person present scenes, $\beta = 0.106$, $SE = 0.012$, $t = 8.70$, and significantly longer in the incongruent gaze cue condition compared to the congruent gaze cue condition, $\beta = 0.031$, $SE = 0.008$, $t = 3.76$. Whether the instruction given to participants made reference to person presence or not had no effect on the time to first fixation ($t < 0.5$), nor was there any difference in time to first fixation between the helpful and unhelpful instruction conditions (t < 1).

Further analyses showed that whether instructions given to participants made reference to person presence or not had no impact on the time to first fixation in person absent versus person present scenes (t < 0.5), nor did it affect time to first fixation in incongruent versus congruent gaze cue conditions (t < 0.5). Having a helpful or unhelpful instruction did not affect time to first fixation in person present versus person absent scenes, nor incongruent versus congruent gaze cue conditions (ts < 1).

Response Time

As in time to first fixation, the response time data required logarithmic transformation in order to satisfy model assumptions of normal distribution. The results are displayed in Figure 35 prior to logarithmic transformation, showing the time taken to press the trigger button on the gamepad indicating search had been completed.

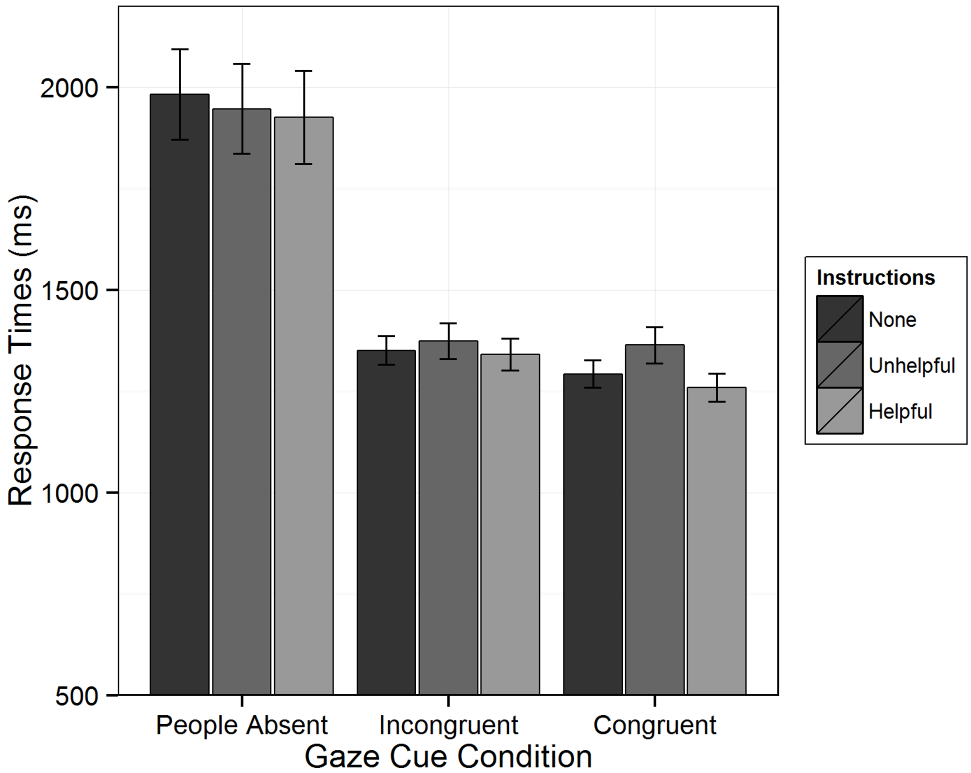

Figure 35. The response time (ms) of the button press indicating location of the target across three gaze cue conditions, where error bars show standard error across all data samples.

Response times were significantly longer in person absent scenes compared to person present scenes, β = 1.318, *SE* = 1.007, *t* = 13.08, and longer in the incongruent gaze cue condition compared to the congruent gaze cue condition, β = 1.604, *SE* = 7.019, *t* = 2.29. While response times were slightly longer when participants were given instruction with no reference to person presence (no instruction) compared to instruction conditions which did make reference to person presence (helpful/unhelpful), β = 1.209, *SE* = 7.675, *t* = 1.57, this was not significant. Similarly, response times in the unhelpful instruction condition were slightly longer than those in the helpful instruction condition, β = 1.468, *SE* = 8.711, *t* = 1.68, but again this was not significant.

Analyses of the interaction between gaze cue condition and instruction condition showed that whether instruction made reference to person presence or not had no impact on response times in person absent versus person present scenes, nor did it affect response times in incongruent versus congruent gaze cue conditions (ts < 0.5). Having a helpful or unhelpful instruction did not affect response times in person present versus person absent scenes, nor incongruent versus congruent gaze cue conditions (ts < 1).

Scan Path Ratio

Although in this model the data were not normally distributed no transformation was performed, as for search efficiency it is to be expected that the majority of responses fall at the lower end of the scale with a lower ratio. The results are presented below in Figure 36.

Analyses showed that the scan path ratio was much higher in person absent scenes compared to person present scenes, $\beta = 0.176$, $SE = 0.063$, $t = 2.79$, and that the congruent gaze cue condition produced more efficient searches than the incongruent gaze cue condition, $\beta = 0.181$, $SE = 0.046$, $t = 3.86$, with scan path ratios closer to one. There was little effect of instruction, with reference to person presence having no effect on participants' scan path ratio compared to no reference to person presence ($t < 1$), and there was no difference in scan path ratio between the helpful and unhelpful instruction conditions, $\beta = 0.067$, $SE = 0.059$, $t = 1.13$.

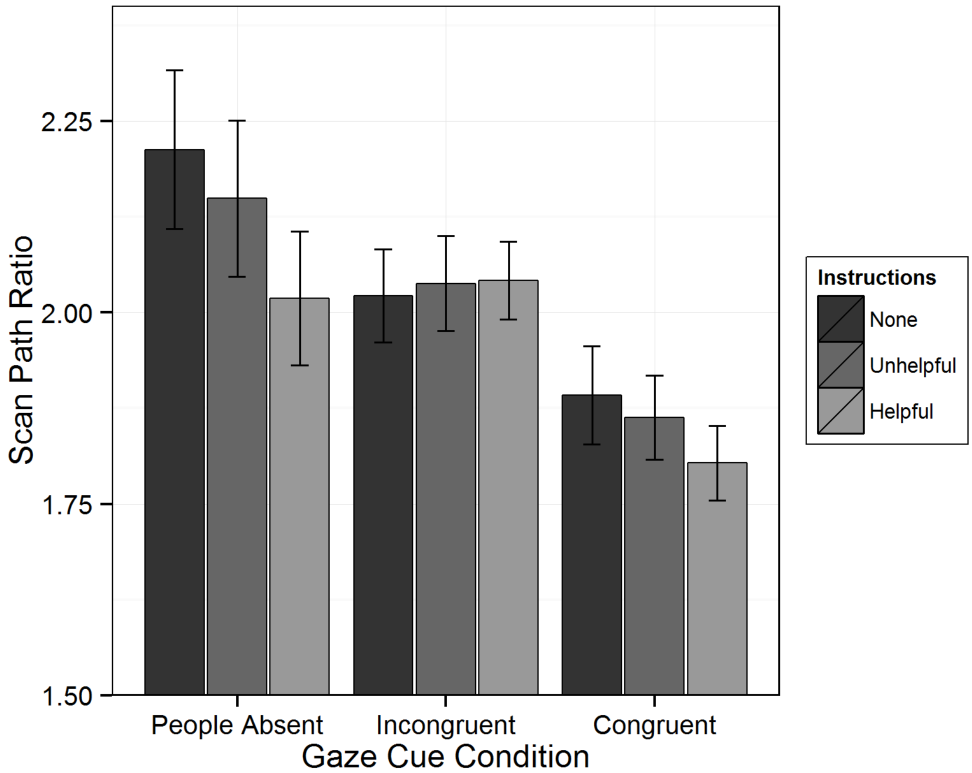

Figure 36. The scan path ratio, indicating search efficiency, across three gaze cue conditions where error bars show standard error across all data samples.

Investigation of the interaction between gaze cue condition and instruction condition showed that whether instruction made reference to person presence or not had no impact on scan path ratio in person absent versus person present scenes, nor did it affect search efficiency in incongruent versus congruent gaze cue conditions (*t*s < 0.5). Having a helpful or unhelpful instruction did not affect scan path ratio in person present versus person absent scenes, nor incongruent versus congruent gaze cue conditions (*t*s < 1).

Error Rate

Figure 37 below shows the proportion of error response trials for each instruction condition across the three gaze cue conditions. An error response was defined as one where a participant pressed the trigger button on the gamepad to indicate they had found the target, without ever fixating it.

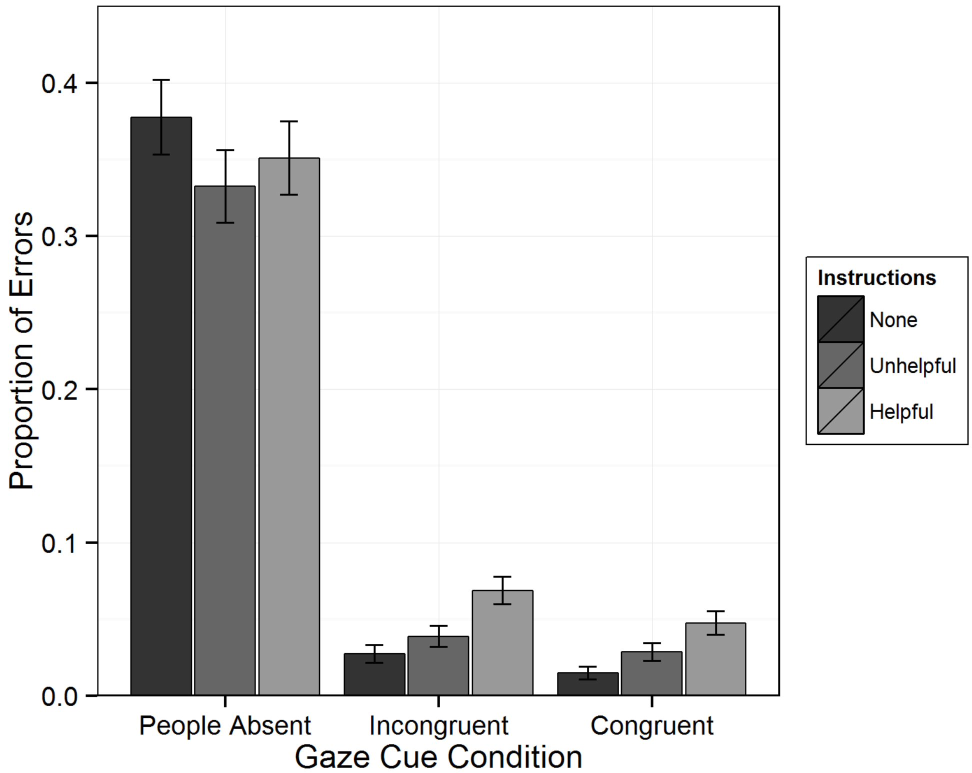

Figure 37. The proportion of errors, with errors defined as a false-positive response in a trial, across three gaze cue conditions where error bars show standard error across all data samples.

There were substantial differences in error rate between person absent and person present scenes, with person absent scenes resulting in a significantly higher error rate, $\beta = 0.319$, $SE = 0.017$, $t = 18.468$. Although there was a slightly higher error rate in the incongruent gaze cue condition compared to the congruent gaze cue condition, $\beta = 0.014$, $SE = 0.008$, $t = 1.800$, this was not significant. Error rate did not differ depending on whether participants had been given instruction that made reference to person presence compared to instructions that did not ($t < 1$), but there were more errors in the helpful instruction condition compared to the unhelpful instruction condition, $\beta = -0.22$, $SE = 0.009$, t -2.479.

Further analyses showed that if the instruction given to participants made reference to person presence, this had a significant effect on performance in the person present versus person absent gaze cue conditions, $\beta = 0.061$, $SE = 0.018$, $t = 3.262$. However, instructions that did or did not make reference to person presence had no effect on error rate in incongruent versus congruent gaze cue conditions ($t < 1$). Whether the instruction was helpful or unhelpful had no bearing on the error rate in person present versus person absent scenes, nor incongruent versus congruent gaze cue scenes (ts < 1). This suggests it is only the fact that the instructions make person presence salient that leads to the increased error rate in person present versus person absent scenes, but this has no further effect once a person is present in the scene.

Overt Gaze-Seeking

An instance of overt gaze-seeking was defined as a trial within which a fixation was made on the person's face at any point in the trial. To investigate any overt gaze-seeking, only person present trials in which participants had made a correct response were used. This is simply because when an error is made, it cannot be determined what search strategy the participant is employing. The analyses here are the first to statistically compare overt gaze-seeking: to do so, the proportion of correct-response trials that featured at least one fixation on the person's face in the person present gaze cue conditions have been compared across each instruction type. The data are presented below in Figure 38. While it is evident in Figure 38 that the proportions of fixations on the face of the person in the scene are never particularly high, there are some clear effects of instruction type on this proportion.

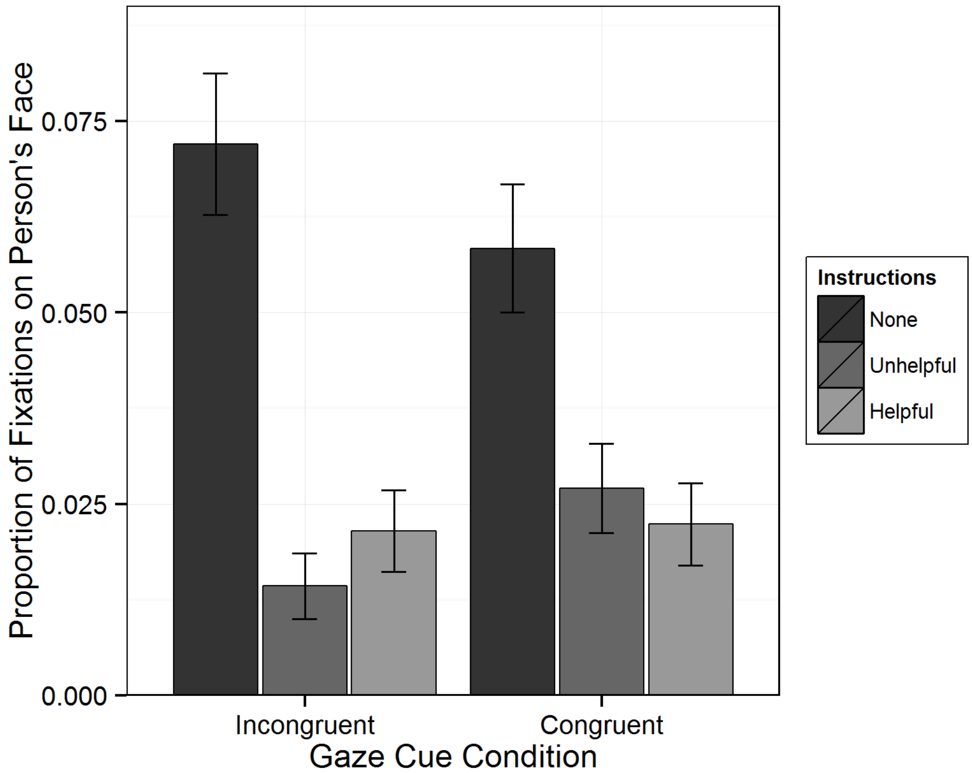

Figure 38. The proportion of fixations on the face of the individual present in the scene across the two person present gaze cue conditions where error bars show standard error across all data samples.

Contrasts of overt gaze-seeking consider only instruction condition, as there was no person absent scenes included for this analysis. Contrasts showed that instructions that did not make reference to person presence resulted in a significantly higher proportion of fixations on the person's face compared to instructions that did make reference to person presence, $\beta = 0.054$, $SE = 0.007$, $t = 6.886$. However, there was no difference in the proportion of fixations on the person's face in the helpful versus the unhelpful instruction condition ($t < 1$).

Further analyses showed that there was a slightly higher proportion of fixations on the person's face in the congruent gaze cue condition compared to the incongruent gaze cue condition depending on whether instructions did or did not make reference to person presence, $\beta = -0.021$, $SE = 0.011$, $t = -1.882$, but this was not significant. There was no difference in performance between the congruent and incongruent gaze cue conditions if a helpful or unhelpful instruction was given to participants ($t < 1$).

Discussion

This chapter has explored the effects of instruction on observer eye movement behaviour in three different gaze cue conditions and across several different measures of performance. Unlike previous chapters, which compare performance across gaze cue conditions, this chapter compares performance across instruction condition within each gaze cue condition in order to provide a quantitative comparison of instruction. To determine whether each instruction condition affects observer eye movement behaviour and if so, in what way, it is necessary to compare the three different studies to one another in an omnibus chapter like this one. Of particular interest in this chapter is the role of the helpful instruction condition as it is the one condition that was formerly unexplored in previous Posner-type tasks. Furthermore, the analyses within this chapter have explored whether there are any interactions between the gaze cue condition presented and the type of instruction given in the task.

Considering first the search initiation phase (comprised of first saccade latency, direction and end point accuracy measures), the majority of results demonstrated that there was no difference in performance elicited by instructions that made no reference to person presence (no instruction condition) and instructions that told participants the person in the scene may or may not be helpful in finding the target object (helpful and unhelpful instruction conditions). The only exception to this was in the measure of first saccade latency. The no instruction condition resulted in shorter first saccade latencies than either the helpful or unhelpful instruction condition. When instruction conditions which made person presence salient were compared, it was found that giving participants an instruction that the person in the scene may be helpful in finding the target actually resulted in longer first saccade latencies than when participants were told to ignore the person. These results suggest that instruction has a considerable effect at the pre-saccadic launch processing stage. The strongest effects of instruction were found in the first saccade latency measure. As search progresses, these effects become slightly weaker in measures of first saccade direction once the eye movement has begun, and disappear almost completely when the first saccade ends, with end point accuracy measures showing very little difference in accuracy between the three instruction types. As search progressed to the scene scanning phase there was little evidence of an effect of instruction with most measures resulting in identical levels of performance across instruction conditions. Effect of instruction only returned, to some degree, at the termination of search, in measures of error, and overall measures of overt gaze-seeking. Error responses, defined as when a participant gives a false-positive response by pressing the trigger button on the gamepad indicating they have found the target without ever fixating it, were significantly greater in number in the helpful instruction condition than the unhelpful instruction condition once a person was present in the scene. Considering these results together, it seems that instruction regarding person presence within the scene has the greatest effect at the very beginning of the visual search, and in overall measures of error and overt gaze-seeking.

While it is unusual for effects to be found only at the very beginning of search, that manipulations can impact search performance as early as in first saccade latency has been documented before in studies of cognitive load. Knapp and Abrams (2012) showed that a verbal cue, as opposed to a pictorial cue, takes longer to process and this increased processing time can affect subsequent visual search, even if participants are given ample time to process the verbal cue. Research has shown that verbal cues, whilst less descriptive than pictorial cues, still offer some guidance in search by informing the internal target template (Schmidt & Zelinsky, 2009; Yang & Zelinsky, 2009). However, using this type of cue adds more cognitive load as participants are required to access their own internal 'database' of objects to build the target template. Research by Solman, Cheyne and Smilek (2011) suggest this may impact even the earliest stages of search. In their study Solman et al. (2011) monitored participants' eye movements while they completed a visual search task within a simple array of geometric shapes. One condition was a simple search task with no additional cognitive load, a second was a memory condition where participants had to retain the image of simple display, and the third condition combined both tasks. In the memory-search condition, which added considerable cognitive load to the search task, participants were first shown a blank screen and then the simple display to be memorized was shown for 500 ms (normally four squares each in a different colour around a central fixation point), followed by the presentation of the search array. Participants were required to find the single square within the array that had either an upward or downward gap in one of its sides, as opposed to a gap on the left or right side featured in the other squares. As they had expected, Solman et al. (2011) found that increased cognitive load, induced by the addition of a memory task, increased response times without affecting search efficiency. This effect had been found before in a study by Woodman, Vogel and Luck (2001), who explained it by stating that memory load does not affect the search process itself, but rather the processes that occur before and after search. Solman et al. (2011) used eye movement data to interpret participant behaviour across three phases of search: pre-search (the time between trial onset and the first saccade); the time between the first saccade launch and first fixation on the target; and the time between first fixation and participant response. As a result, they were able to demonstrate that increased cognitive load does in fact affect all stages of search, including search initiation. The authors interpret this as evidence of cognitive load resulting in impairment to fixation selection processes during the first stage of search. The increased cognitive load created by giving limited information about the target by using a verbal cue in the studies analysed above is comparable to that created by retaining the array in memory in Solman et al.'s (2011) study. This would suggest that in using a verbal cue can impact performance as early in the search process as first saccade latency, and Solman et al.'s (2011) study provides corroborating evidence of these effects occurring at the earliest stage of search.

However, verbal cues were used in all gaze cue and instruction conditions, which means that this cognitive load would affect all conditions equally. Therefore, the difference in performance across instruction conditions must be accounted for by something else. It has been suggested that the very first part of search is affected by changing the observer's expectation of where the target will be located (Spotorno et al., 2014). Castelhano and Henderson (2007) argue this could be due to how an initial glimpse shapes the way in which observers process a scene. They showed participants digitised photographs of real-world scenes, providing an initial glimpse of 250 ms, followed by the presentation of the name of the target object and then the second presentation of the scene. The scene preview was manipulated so that it would be helpful in directing later eye movements in only some conditions. When given an identical preview there was a clear scene preview benefit, suggesting the initial glimpse provided sufficient information to guide subsequent eye movements. This was not replicated when the authors provided a scene preview that presented the conceptual identity of the scene without specific visual details of the target. Castelhano and Henderson (2007) concluded the internal scene representation generated during the initial glimpse was used to guide later eye movements during visual search.

There are several ways in which Castelhano and Henderson’s (2007) findings can contribute to understanding the impact of instruction in the initial stage of visual search found in the cross-study comparison. Their results supported Spotorno et al.’s (2014) hypothesis that initial glimpses of a scene inform the observer’s global internal representation of the scene. The specificities of whether the target is present or not does not impact this global representation, instead the purpose of this initial processing is to provide the observer with an understanding of likely areas in which the target will appear. In stimuli used in the experiments discussed in this chapter, that initial glimpse can identify the bottom half of the scene where the table is located as the priority for subsequent eye movements searching for the target.

This has been supported by later research conducted by Vō and Schneider (2010), who showed that previewing an object in isolation does not offer any benefit to later search. They presented observers with 3D rendered images of realistic scenes, which the observers had to search for an embedded target object through a gaze-contingent window. Participants received one of four scene previews, which were either identical to the later scene, showed only the scene background, showed only the objects within the scene, or presented a meaningless pixelated control. Participants exhibited the greatest search benefit following identical and background scene previews. Vō and Schneider (2010) suggest that given this evidence, what is crucial for scene preview benefit is the scene as a whole – particularly the background context. This allows the global processing of the scene creating an internal representation of the scene's spatial layout, and combining this information with knowledge about the task restricts the probable areas in which the target will appear.

Considering this evidence in terms of the analyses reported above, the reasoning of Castelhano and Henderson (2007), and Vō and Schneider (2010), would suggest that the initial glimpse of the scene – affecting the first saccade latency measure – guides subsequent eye movements during the search. The global context of person present scenes guides eye movements to the table in the bottom half of the scene as the most probable location for the target object to appear. In person absent scenes, where the contextual information provided by the person's position in the scene is unavailable, observers have nothing to guide them to the table in the bottom half of the scene when searching for the target. First saccade latency is shorter in the unhelpful instruction condition than in the helpful instruction condition, which may suggest that participants disregard gaze as adding to the global context of the in the unhelpful instruction condition. Instead they use only the general background information available (i.e. there is a table in front of a wall, and the observers' understanding of the context would suggest objects would be most likely to appear on the table). A helpful instruction suggests the person, and perhaps the cues they provide, are important for understanding the context of the scene and predicting target location. Processing this additional information could account for the increase in latency prior to launching the first saccade. This may mean that participants can select which information to attend in order to guide their search, depending on the information they have been given about the task.

This point is also pertinent for consideration of the error rate analysis. Results showed that participants were significantly more likely to make a false-positive error response in the helpful instruction condition in person present scenes. If, as suggested above, participants can use task instruction to guide eye movements, and thus allocate attention to the spatial area cued by gaze, they are quite possibly wasting cognitive resources. The cues presented to participants are spatially uninformative (as in previous Posner-type tasks, e.g. Driver et al., 1999; Ricciardelli et al., 2002), which means that they are equally likely to cue the distractor object as they are to cue the target. While this has not elicited the longer response times usually found in Posner-type task research as evidence of the cost of gaze following when this is not beneficial to the task, error rate was not generally reported in these studies. Considering the main goal of the task is to find the target quickly, if a participant attributes greater importance to gaze information based on the task instruction, they are more likely to produce more errors as a result.

Interactions between gaze cue condition and instruction condition were only found in the last stage of search – error rate. The interaction between person-salient instructions (i.e. helpful/unhelpful) and person absent versus person present scenes suggests that mentioning the person in the scene during the task instruction – whatever that instruction is – will cause a significant difference in the proportion of errors made in person absent trials compared to person present trials. However, there is no difference in errors in person absent versus person present trials when looking across the helpful and unhelpful instruction conditions, or across person present scenes. This would suggest that it is the simple mention of a person the leads to deterioration in performance when that person is not present in the scene (i.e. person absent trials). Considering previous research which has documented our clear preference for people over other stimuli (e.g. Birmingham et al. 2008, 2009; Fletcher-Watson et al., 2008; Leder et al., 2010; Zwickel & Vō, 2010), it could be reasoned that this high-level stimuli diverts part of the observer's attention to the task, thus leading to a greater proportion of error when the expected stimulus – a person – is not present. Indeed, Summerfield and Egner (2009) described the necessity for the brain to compensate for the volume of visual stimuli, which cannot all be processed due to limited computational capacity, and that one mechanism by which the brain can select what to process is by prioritising stimuli by expectations of what is likely to appear in our immediate environment. They give the example of walking into a familiar room; elements of the environment that are constant (e.g. the colour and texture of the wallpaper, the shade of the carpet) do not need to be processed in depth on each viewing. This frees computational capacity to process anything new or changed within the environment. However, expectations also facilitate visual processing by providing context information to help identify ambiguous stimuli. For example, a box-shaped object in a kitchen would – by expectations generated from prior knowledge – be more likely to be a bread bin than a file box or per carrier. If, as Summerfield and Enger (2009) suggest, these types of expectation can shape the way in which we perceive visual stimuli, then this may account for the interaction described in the error rate data. If expectations guide visual processing, task instructions that make reference to person presence would create an expectation that a person will be present in the scene. When a person present scene appears, the expectation is met and visual processing can be guided accordingly. However, when a person absent scene appears, the expectation is not met and the observer must adopt a new strategy to process the scene. It is logical then that in this combination – person-salient instruction and person absent scenes – a higher error rate would be anticipated.

However, if it is true that the violation of expected stimuli – in the case of the current research, the presence of a person in the scene – causes a higher rate of errors, why is the same effect not apparent across other performance measures? The measure which provides the strongest evidence of an effect of instruction is the measure of the proportion of fixations on the person's face in the scene (the measure of overt gaze-seeking). Looking only at person present gaze cue conditions, there is a marked increase in the proportion of fixations on the face when participants are given no instruction compared to when they are given either a helpful or unhelpful instruction regarding person presence. If instructions made person presence more salient, regardless of the content of the instruction, the higher proportion of fixations would be expected in the helpful and/or unhelpful instruction conditions. However, it seems that when participants are not given any information regarding the person, they are more likely to fixate on their face.

One explanation for this counter-intuitive result comes from evidence that suggests when task information is incomplete participants will utilise gaze cues in an attempt to extract more information that is useful to the task. This was discussed by Macdonald and Tatler (2013a), who conducted a mobile eye tracking study to determine how participants used gaze cues to disambiguate instructions given by the experimenter. Participants were instructed to select blocks from an array in front of them, and the correct block was indicated by an instruction that was either ambiguous or unambiguous, and was either accompanied by a gaze cue or by no gaze cue. This was a unique manipulation where gaze only added helpful information when the verbal instruction was ambiguous, giving Macdonald and Tatler (2013a) the opportunity to measure to what extent participants utilised gaze cues when they were helpful and unhelpful. When participants were given an unambiguous language cue and a helpful gaze cue they very rarely made fixations on the experimenter's face. However, when given an ambiguous language cue and a helpful gaze cue, there were significantly more fixations on the experimenter's face.

Macdonald and Tatler (2013a) suggest that their study provides evidence that supports the theory that observers restrict fixations to task-relevant stimuli, and that social cues – including gaze cues – can fall into this category. This may provide an evidential basis for the high proportion of fixations on the person's face in the no instruction condition, where participants may use gaze to add crucial information to their target template (as discussed by Malcolm & Henderson, 2009), and why this proportion falls when participants are given a unhelpful instruction that informs them the person present is not relevant to the task. However, it does not account for that same low proportion in the helpful instruction condition. If observers direct their fixations based on task relevance, when they are informed the person may help them find the target object faster, it seems logical this makes that cue task relevant. Yet, the corresponding increase in fixations is not apparent, and it is unclear why. It is possible that the incidence of overt gaze-seeking is simply too low for statistical differences to emerge. It may also suggest there are problems with the instruction manipulation.

Studies that have examined the effects of task instruction have typically shown considerable differences in observer eye movements depending on what instruction condition they were in. This is true across real world and laboratory-based studies.

In real world studies where participants are asked to complete an everyday task such as making a cup of tea (Land et al., 1999) or a sandwich (Hayhoe et al., 2003), fixations are generally limited to whatever is task relevant, with gaze moving slightly ahead of the hands. When participants are instead asked to view something without interacting with it, their eye movements change depending on what they are asked to assess – in Yarbus's (1967) seminal work it was demonstrated that people would view the same painting in many different ways according to what they were assessing (e.g. wealth, age of subjects, etc.) Richardson, Hoover and Ghane (2008) and Richardson et al. (2012) demonstrated that changing task instructions to manipulate participants' belief of whether they were viewing an image alone, or that the person in the next room was looking at the same pictures, changed their eye movements as they viewed the scene.

Even laboratory-based studies have evidenced that participants modify their fixations depending on task – Itier et al. (2007) showed participants would fixate on a face differently depending on whether they were discriminating eye or head direction, and Fletcher-Watson et al. (2008) demonstrated participants would fixate more frequently on a subject's torso when asked to discriminate gender. Why then, when the effects of instruction seem robust across very different methodologies, have the same effects not been encountered in this analysis?

It is possible that the type of task assigned in the current study makes it harder to manipulate the perceived helpfulness of the provided gaze cue. For example, in the studies cited above the differences in instruction directly related to the participants' task. Participants were asked to discriminate something specific about the scene, or alternatively to discriminate another feature or to freely view the scene (e.g. Itier et al., 2007; Fletcher-Watson et al., 2008). They were asked to assess different features of a painting (Yarbus, 1967), or perform a different everyday task (Hayhoe et al., 2003; Land et al., 1999). However, in the current studies participants were always asked to complete the same task: to find the target object. The information regarding person presence and the helpfulness of their gaze cue was supplementary and it is quite possible participants disregarded the information. Haider and Frensch (1999) tested an information reduction hypothesis that suggested participants could voluntarily select which elements of task information they would follow by asking participants to perform a verification task whilst prioritising either speed or accuracy, or both (one after the other). Their participants changed the information they attended based on what they had to prioritise, which would support the hypothesis that participants in the current study focused on the main task (find the target object) and may have disregarded additional information concerning the person in the scene. Alternatively, participants may have not believed the manipulation. This is an inherent problem when using a participant sample mostly comprised of undergraduate Psychology students, who are aware of manipulations occurring in the experiments in which they participate, and has been highlighted as an issue on many occasions (e.g. Cannon, Higginbotham, & Leung, 1988; Podsakoff, MacKenzie, Lee & Podsakoff, 2003; Schultz, 1969; Smart, 1966). That the performance in helpful and unhelpful instruction conditions is so consistent across the majority of measures would certainly support this assumption. It may also be the case that any instruction that makes person presence salient, regardless of its content, elicits similar effects on participants' eye movements. This is speculative, based on the established informational hierarchy that places information from people higher than other stimuli (see Birmingham et al., 2009), and cannot be confirmed nor denied from evidence discussed within this chapter, but is a possible explanation future research could look to explore.

This chapter presents a quantitative analysis of the effects of instruction on participants' eye movements in a single-cue visual search task. Participants were either given no instruction regarding person presence in the scene, were told that the gaze cues were unhelpful and to be ignored, or that the cues may be helpful in finding the target object. The analyses demonstrate that instruction seemed to have little effect across the majority of performance measures, but did however elicit different first saccade latencies and error rates depending on the instruction given to participants. While there was an interaction between person-salient instruction and error rate in person absent versus person present scenes, there were no other interactions between gaze cue condition and instruction condition in any of the other performance measures. There was apparent variation in the degree of overt gaze-seeking depending on the instruction condition, with the highest rate of fixations on the person's face in the no instruction condition. It has been suggested that participants may be attempting to disambiguate the scene context by utilising gaze in this scenario, however it is difficult to account for all the gaze-seeking evidence with this explanation. Generally the performance in helpful and unhelpful instruction conditions seldom differ, which may suggest this manipulation was not entirely successful. This may be due to a number of factors, but perhaps most likely because the person-presence information was supplementary to actual task instruction and as such may have been disregarded by participants. Cumulatively this evidence suggests that, if the instruction manipulation was successful, participants' derive a benefit in the visual search task from person presence, but what they are told about the purpose of person-presence has no effect on how they use gaze information.

However, based on the evidence discussed here, it is not possible to be entirely confident in the success of the manipulation, and as such these findings should be interpreted tentatively.

Chapter Five

Examining how multiple simultaneously-presented gaze cues impact observer eye movement behaviour in visual search

Introduction

The previous experimental chapters have explored a novel Posner-type paradigm that examines the effects of a social gaze cue when presented in a realistic social scene. These studies have demonstrated that when presented with more realistic stimuli, not all of the eye movement behaviour observed matches with what would be predicted from previous laboratory-based paradigms. By testing the paradigm across several different types of instruction and eliciting the same benefits of person presence across all three, a degree of confidence can be had in the robustness of these effects. These chapters represent a solid foundation of evidence for using this kind of more realistic stimuli within a visual search task, and provide a baseline from which other phenomena can begin to be explored.

The focus of this chapter is on extending the confines of this new paradigm to come even closer to the sorts of social gaze cues we receive in the real world. In our everyday environment social gaze cues rarely occur in isolation. They are instead presented with the full contextual information of head and body movement, and in most typical daily scenarios we encounter more than one person providing gaze cues at a time. If we were to imagine walking through a busy shopping centre and navigating through the crowd, even the anecdotal evidence from these experiences demonstrate that more often than not we are faced with more than one set of gaze cues at a time. However, there is currently no empirically controlled study of simultaneously presented multiple gaze cues. Posner-type tasks that use only one person to provide gaze cues suggest that orienting to another person's gaze cue is a reflexive behaviour that is very difficult to inhibit (e.g. Langton & Bruce, 1999; Ricciardelli et al., 2002). Yet, considering the example of walking through a busy shopping centre, we do not follow each individual's gaze as it is presented to us. The question then is how we respond to these simultaneously presented multiple gaze cues.

Presently, the only research that examines the presentation of multiple gaze cues comes from real world research, and the results from these studies are quite different from what would be expected given the evidence from laboratory-based research. Gallup, Chong and Couzin (2012a) conducted an experiment where a shiny attractive object was placed in the centre of a busy corridor. This object was constructed of one-way reflective Plexi-glass, behind which cameras were placed to record the directed looks of passers-by. The authors wished to determine to what extent pedestrians would spontaneously engage in joint attention on the attractive object, which would involve following the gaze of an oncoming pedestrian. What they found was in stark contrast to the predictions of reflexive gaze following that comes from the traditional body of research on gaze following. Instead, Gallup et al. (2012a) found that pedestrians were *less* likely to look at the object if an oncoming pedestrian had looked at it. They were more likely to look at the object at the same time as someone else if that person was travelling in the same direction as them, which is therefore not an eye gaze following behaviour, but a following of head direction. The authors hypothesised that this was an example of our unwillingness to engage in joint attention with a stranger, and that instead we modulate our gaze behaviour depending on the present social context.

The disparity between the findings of laboratory-based and real world studies raise the question of whether these types of experiments can truly be compared. In the controlled setting of the laboratory, ecological validity can be lost, which has already been discussed in previous chapters. However, real world research does not come without its drawbacks, as without some element of control over the variables studied it is impossible to be sure that the intended effect of the independent variable is being measured. In the present study the effects of multiple gaze cues on viewing behaviour were studied using photographic images rather than real world situations. Presenting observers with two gaze cues simultaneously more accurately reflects the types of social gaze cue encountered in everyday life and using photographs of real people giving a natural gaze cue is more ecologically valid than previous Posner-type task research (e.g. Friesen & Kingstone, 1998; Driver et al., 1999), but allows control over the direction of these gaze cues and minimises external confounding variables.

The present study considers the influence of multiple simultaneously present gaze cues on the search behaviour of observers viewing photographs of real world scenes. Once again these stimuli draw upon the key concepts of traditional Posner-type tasks, using non-predictive cues from a centralised position to either the left or right side of the scene. These cues are provided in a realistic context: they involve the movement of eyes and head, and are shown within the context of the upper half of the individuals' bodies as they sit behind the table. As in previous chapters that explored a single cue variation of this paradigm, the cues are not shown in the centre of the screen; instead an eye movement must be made to overtly seek out gaze cues. This allows the examination of observer eye movement behaviour within a controlled setting that more accurately reflects real world encounters. Having multiple gaze cues also allows for an additional gaze cue condition. As well as congruent and incongruent cues, where both people cue the same object, the current study also presents participants with conflicting gaze cues, where each person cues a different object. The types of gaze cues presented in this study are analogous to those we might experience in the example of navigating a busy shopping centre, and it is only by having multiple gaze cues within the scene that this more realistic range of gaze cues can be tested.

Method

Participants

A total of 62 people (20 male) were recruited for participation in this study. All had normal or corrected vision and were naïve to the purposes of the study. Level one and two undergraduate students received course credits for participation; anyone not eligible for course credit was paid £2. None of the participants had taken part in any of the previous studies.

Materials

Experimental scenes were created using ten different sets of everyday objects. Each scene featured one of the ten sets of 15 everyday items arranged on a table top. Within each scene one item was designated the target and another designated the distractor. These items were always on the opposite sides of scene centre, so that looks to these items are unlikely to be due to the typical human bias to look near the scene centre irrespective of content (as discussed in Tatler, 2007). The target was equally likely to appear on the left or right third of the table, and would never appear in the centre. An example of one of these arrangements is shown in Figure 39 below.

Figure 39. Examples of each gaze condition for one arrangement of objects on the table top. In this case the target item was the Dr Who DVD box set, and the distractor was the tub of hot chocolate.

In Figure 39, Box A indicates the *congruent* gaze cue condition, where both people cue the target. Box B shows the *incongruent* condition, where both people cue the distractor. Boxes C and D show the *conflicting* gaze cue condition, where one person cues the target and the other cues the distractor. Two versions of the conflicting gaze cue condition were created such that the individual cueing the target could be counterbalanced within and across participants, and so that there would be conflicting gaze cues where each person looked to the opposite side of the table from which they were sitting, and at the same side of the table as where they were sitting.

There were a large number of factors which required counterbalancing within this study. Firstly, the position of the target was counterbalanced to appear on the left third of the table in the scene in half of the trials, and on the right third of the table in the other half of trials. The position of the people in the scene was also counterbalanced so that each person appeared on the left for an equal number of trials as they appeared on the right. Furthermore, scenes presented to participants were counterbalanced to ensure that each person cued the target as many times as they cued the distractor, and that each person cued the target from the left of the scene as many times as they cued from the right of the scene.

Full counterbalancing required a large number of experimental scenes, and creating object arrays of 15 different objects for each one would have been exceedingly difficult. For this reason, it was decided multiple arrangements would be used to allow repeated use of the same object sets, provided each arrangement used different target and distractor objects (which were randomly selected) to prevent any learning effects. In the first arrangement the objects would be arranged on the table with the two selected items that would either be used as a target or distractor – depending on which item the participant was asked to look for – on opposite sides of the table. This arrangement would be photographed four times, as shown in Figure 39. The objects would then be rearranged, with different items positioned as target and distractor, and the four gaze cue conditions would be photographed again. This was repeated two more times, creating four sets of people present scenes for each set of objects. A final shot would be taken with the objects rearranged once more, but with no people present in the scene. In total 10 people absent scenes and 160 person present scenes were created. Eight versions of the experiment were made so that only one of the four photographs for each arrangement paired with either the first or second target would be seen by one participant.

Eye Tracking

Eye movements were tracked using the same equipment as stated in previous chapters. The same procedure was followed to accurately calibrate the desk-mounted eye tracker, and the same acceptance criteria (average spatial error less than 0.5 degrees and maximum error less than 1 degree over the 9 calibration points) were used.

Procedure

Trials followed the same procedure as in previous chapters. To recap: a single-point calibration check was performed before each trial began. The name of the target object was presented for 500 ms, followed by a blank screen for 500 ms, and then the presentation of the scene for a maximum of 10 s, or until the trigger button on the gamepad was pressed by the participant indicating they had found the target object. The number of scenes in this study differs slightly from the others, as this was the first study tested using the Experiment Builder software. Each participant saw a total of 50 scenes: 10 people absent scenes and 40 person present scenes. In all eight versions of the experiment, all ten people absent scenes would be presented. A total of 40 person present scenes were each shown paired with either the first or second randomly-selected target.

Participants were given no instruction regarding the presence or absence of a person in the scenes. They were given a brief description of what would happen in each trial and simply asked to find the target object as quickly as possible.

Data Analysis

Following the structure of analyses in previous chapters, data were analysed using linear mixed effects models (LMMs) in the lme4 package (Bates et al., 2011) within the R statistical analysis environment (R Development Core Team, 2011). In linear models, the lmer() function returns t-values without the associated p-values, so any effects for which the t-value is greater than two – that is effects larger than twice their standard error – as reflecting a significant effect (as in Kleigl et al., 2012). In the analyses below, it is noted if any data required logarithmic transformation to generate a normal distribution, and the model with the most complicated random effects structure that converged was reported. In the following analyses, measures of search initiation (first saccade latency, direction, and end point accuracy) consider data from all trials, irrespective of whether the participant fixated the target, as did the measure of error rate. The time to first fixation, response time, scan path ratio, and overt gaze-seeking analyses used only correct-response trials, where participants had both fixated the target and pressed the trigger button to indicate they had found it.

Results

Search Initiation

As in previous chapters, the first stage of search is comprised of first saccade latency, the direction of the first saccade, and the end point accuracy of the first saccade. First saccade latency analysis required two transformations. The data included a small number of very short latencies, which most likely were the results of pre-emptive eye movements beginning before the appearance of the scene. Very short latencies were defined as being less than 100 ms. A total of 390 trials featured a very short latency (12.58% of the total number of trials), and these very short latencies were removed. The remaining data underwent logarithmic transformation to generate a normal distribution. These data are presented in their original form in Figure 40.

When compared to the people absent condition, both the incongruent, β = -0.013, *SE* = 0.005, *t* = -2.37, and conflicting gaze cue conditions, β = -0.011, *SE* = 0.005, *t* = -2.25, resulted in significantly shorter first saccade latencies. While there was some reduction in first saccade latency in the congruent condition when compared to the people absent condition, this was not significant, β = -0.007, *SE* = 0.006, *t* = -1.23. When comparing between the three people present gaze cue conditions, there were no significant differences in first saccade latency between any of the conditions (*t*s < 1).

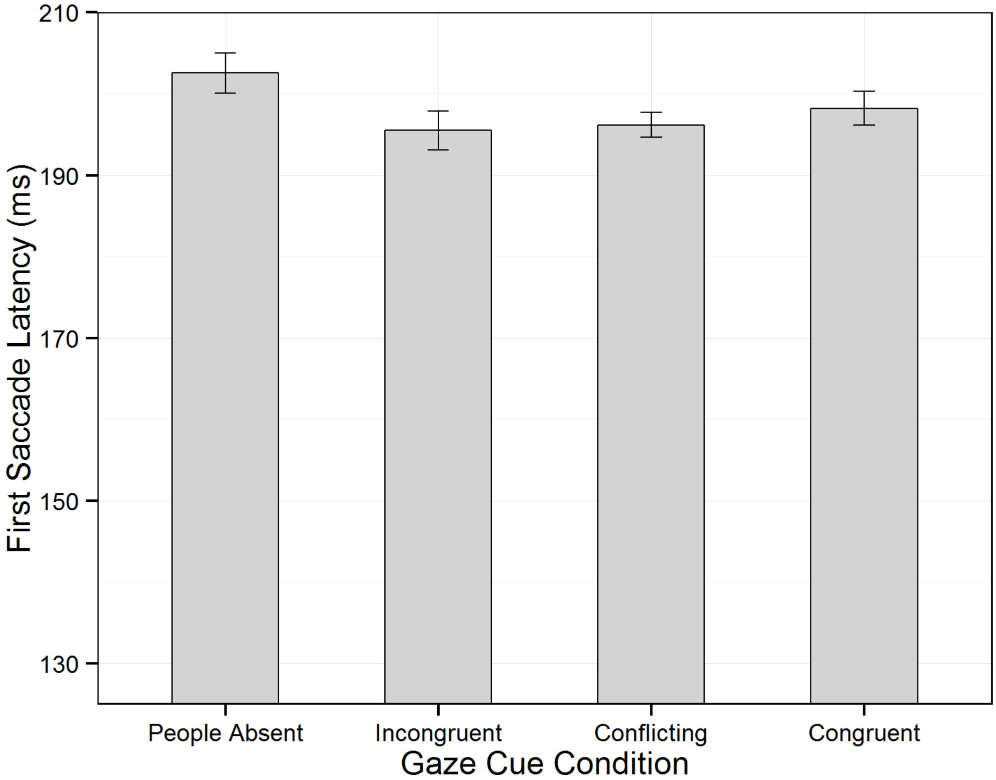

Figure 40. First saccade latency in each of the four gaze cue conditions. Error bars show standard error across all data samples.

The proportion of first saccades directed toward the target showed a greater degree of variation across the four gaze cue conditions, as can be seen in Figure 41. In this measure the people absent and incongruent gaze cue conditions displayed similar proportions of first saccades directed toward the target ($p > 0.05$). However, if at least one gaze cue was directed toward the target an increase in first saccades directed toward the target was seen. Both the conflicting, $\beta = 0.356$, $SE = 0.115$, $z = 3.088$, $p = 0.002$, and congruent gaze cue conditions, $\beta = 0.552$, $SE = 0.151$, $z = 3.647$, $p < 0.001$, resulted in a significantly higher proportion of first saccades directed toward the target than the people absent gaze cue condition.

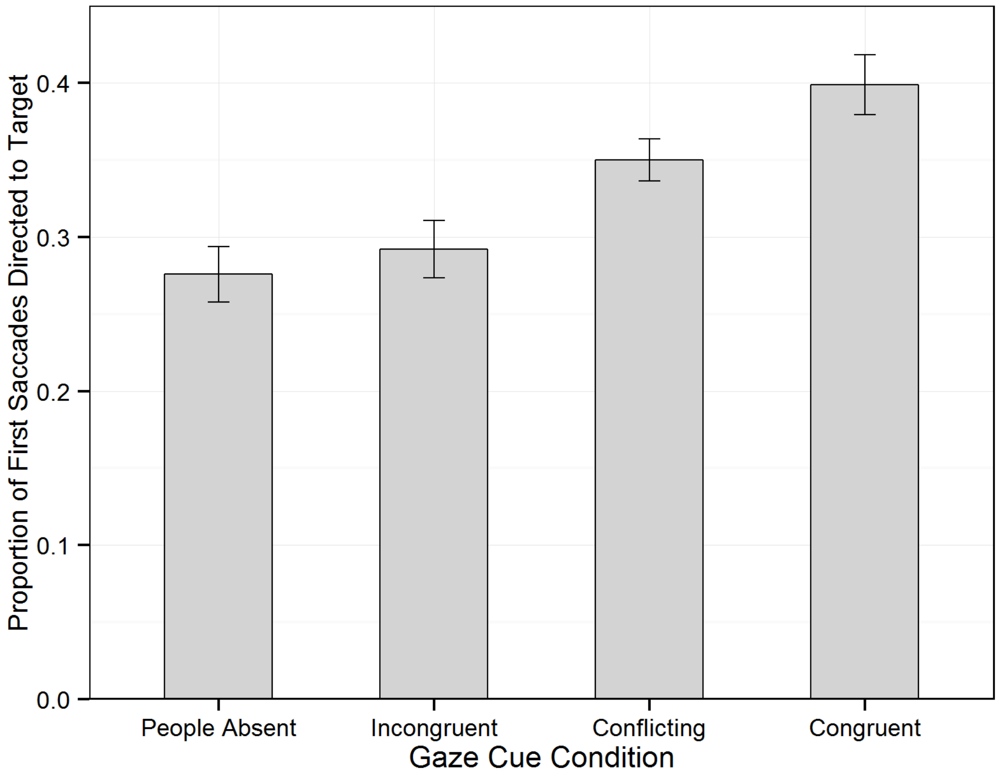

Figure 41. The proportion of first saccades directed toward the target across four gaze cue conditions. Error bars show standard error across all data samples.

Comparing between people present gaze cue conditions showed that the incongruent gaze cue condition produced a lower proportion of first saccades directed toward the target than both the conflicting, $\beta = 0.278$, $SE = 0.130$, $z = 2.132$, $p < 0.05$, and congruent, $\beta = -0.474$, $SE = 0.133$, $z = -3.545$, $p < 0.001$, gaze cue conditions. There was no difference in the proportion of first saccades directed toward the target between the conflicting and congruent gaze cue conditions ($p > 0.1$).

The final measure for search initiation was the end point accuracy of the first saccade. As can be seen in Figure 42, the people absent gaze cue condition brought the eyes closer to the target than any of the people present gaze cue conditions.

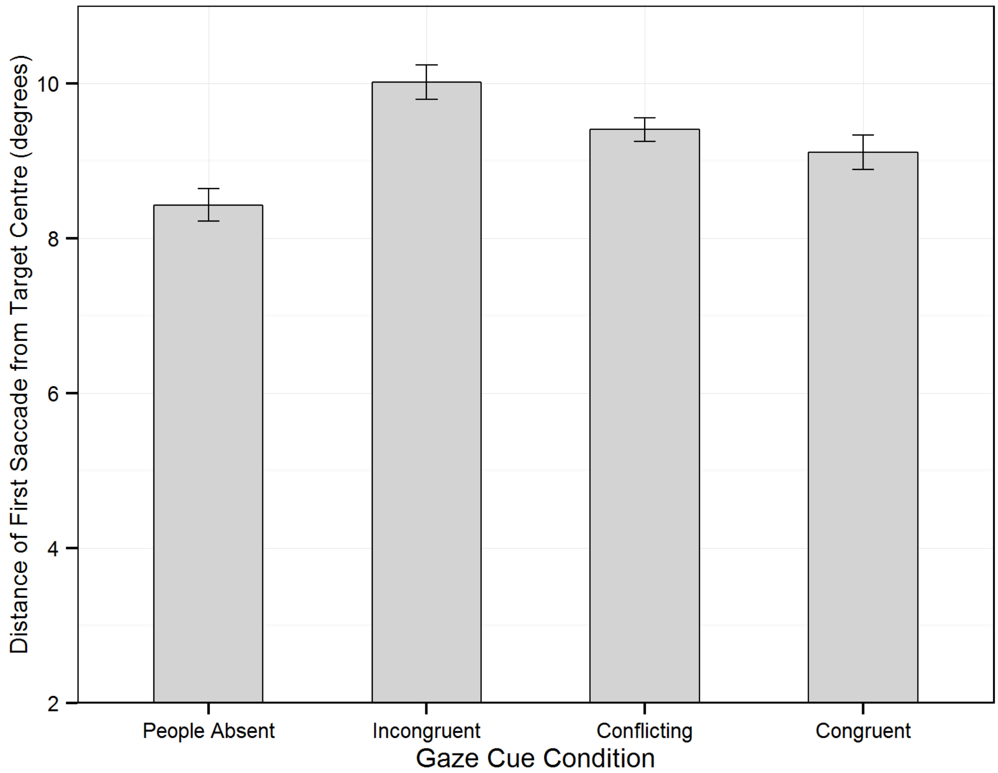

Figure 42. The distance of the landing point of the first saccade from the centre of the target ROI (in degrees of visual angle) as a measure of end point accuracy across four gaze cue conditions. Error bars show standard error across all data samples.

Analysis showed that the people absent gaze cue condition resulted in significantly more accurate first saccades than the incongruent, β = 1.541, *SE* = 0.298, *t* = 5.161, conflicting, β = 0.969, *SE* = 0.256, *t* = 3.778, and congruent gaze cue conditions, β = 0.672, *SE* = 0.323, *t* = 2.080. The incongruent condition was the least accurate of the three people present gaze cue conditions, resulting in less accurate first saccades than both the conflicting, β = -0.571, *SE* = 0.266, *t* = -2.146, and the congruent gaze cue condition, β = -0.866, *SE* = 0.319, *t* = -2.712. While there was some improvement in accuracy in the congruent gaze cue condition as compared to the conflicting condition, β = 0.284, *SE* = 0.256, *t* = 1.108, this was not significant.

Scene Scanning

Again following the same structure as in previous experimental chapters, the scene scanning stage of search comprises of analyses of the time taken to first fixate on the target, overall response time, scan path ratio, error rate, and analysis of overt gaze-seeking.

People presence had a very strong effect on the time taken to first fixate on the target, a measure that required logarithmic transformation for analysis. All three people present gaze cue conditions resulted in earlier first fixations than the people absent condition. Although the incongruent gaze cue condition was the closest of the three to the people absent condition, β = -0.048, *SE* = 0.015, *t* = -3.03, there was still a clear decrease in the time taken for participants to first fixate the target than when there were no people present in the scene. The time to first fixation increased slightly when one gaze cue was directed toward the target in the conflicting gaze cue condition, β = -0.046, *SE* = 0.013, *t* = -3.39, and to an even greater degree when both gaze cues were directed toward the target in the congruent gaze cue condition, β = -0.089, *SE* = 0.015, *t* = -5.63. These data are presented in Figure 43.

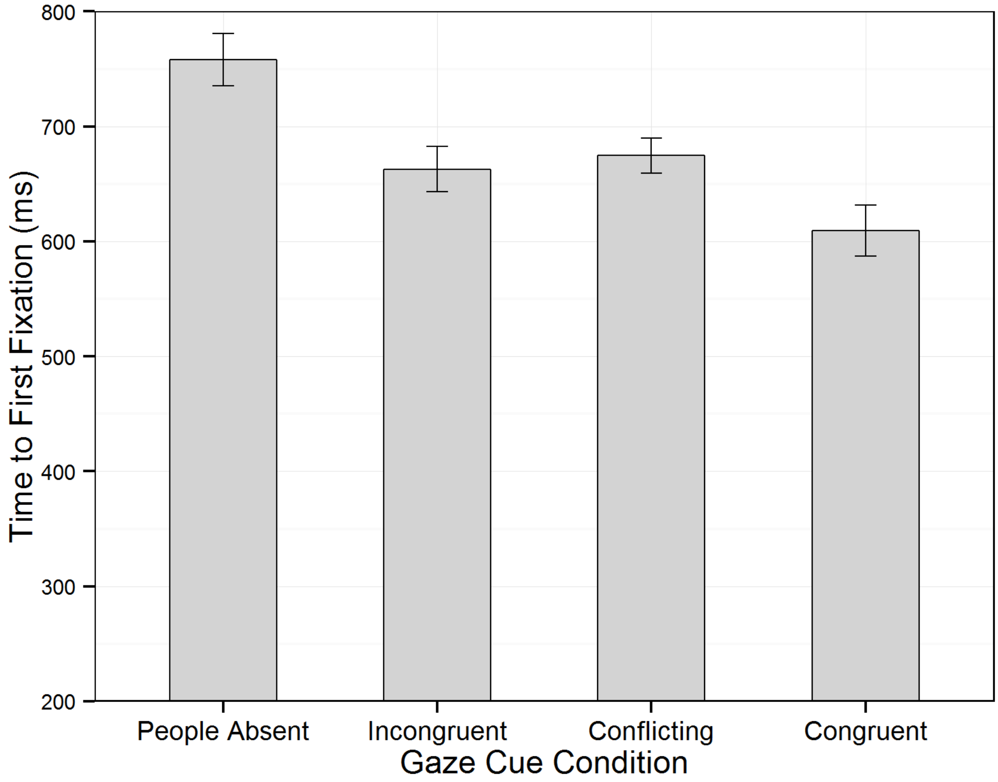

Figure 43. The time to first fixation on the target (ms) from scene presentation across four gaze cue conditions. Error bars show standard error across all data samples.

As might be expected from Figure 43, the incongruent and conflicting gaze cue conditions resulted in highly similar times to first fixate the target (t < 0.5). In comparison, the congruent gaze cue condition resulted in first fixations on the target that occurred significantly earlier than both the incongruent, β = 0.041, *SE* = 0.016, t = 2.52, and the conflicting, β = 0.043, *SE* = 0.013, t = 3.09, gaze cue conditions.

Like the time to first fixation on the target, response times required a logarithmic transformation to satisfy model assumptions. As can be seen in Figure 44, people presence resulted in considerably faster response times.

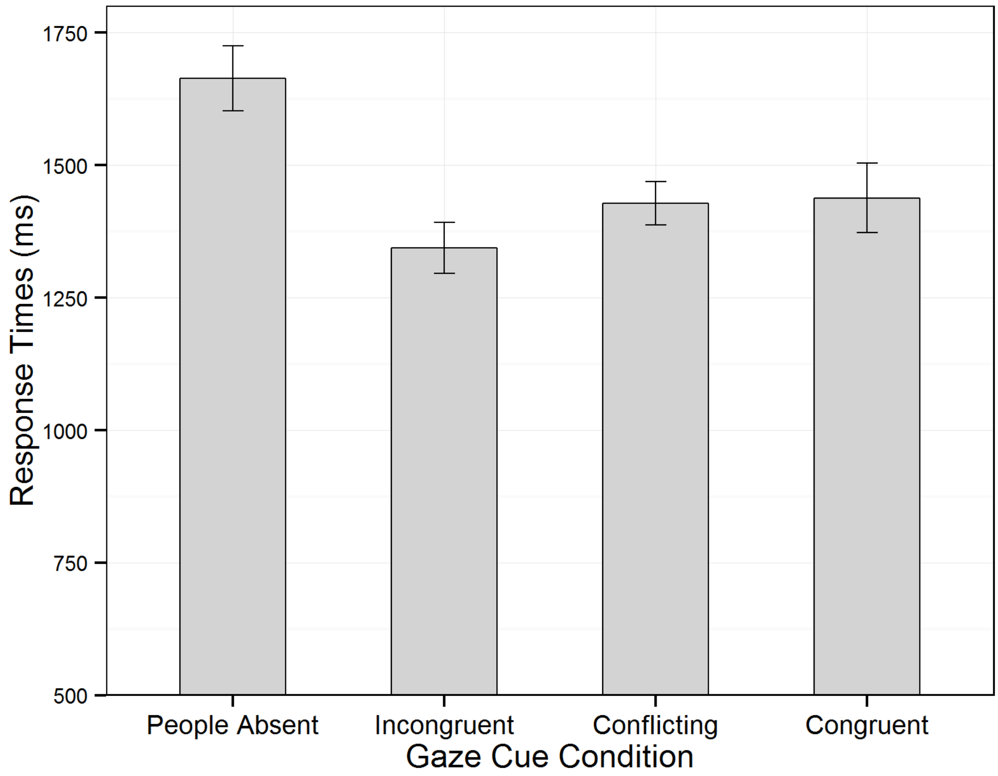

Figure 44. Response times (ms) to button press indicating successful search for the target across four gaze cue conditions. Error bars show standard error across all data samples.

When compared to the people absent gaze cue condition, the incongruent, β = -0.076, *SE* = 0.013, *t* = -5.53, conflicting, β = -0.067, *SE* = 0.012, *t* = -5.54, and congruent, β = -0.075, *SE* = 0.013, *t* = -5.43, gaze cue conditions all resulted in significantly less time taken for participants to indicate they had located the target. Further analysis comparing the three people present gaze cue conditions to each other showed no differences between the different conditions (*t*s < 1).

Considering search efficiency, an optimal route would be a single saccade from the fixation point at scene onset to the centre of the target object; therefore more efficient searches should have a scan path ratio closer to 1. Although in this model the data were not normally distributed no transformation was performed as for search efficiency it is to be expected that the majority of responses fall at the lower end of the scale with a lower ratio. The results are presented below in Figure 45.

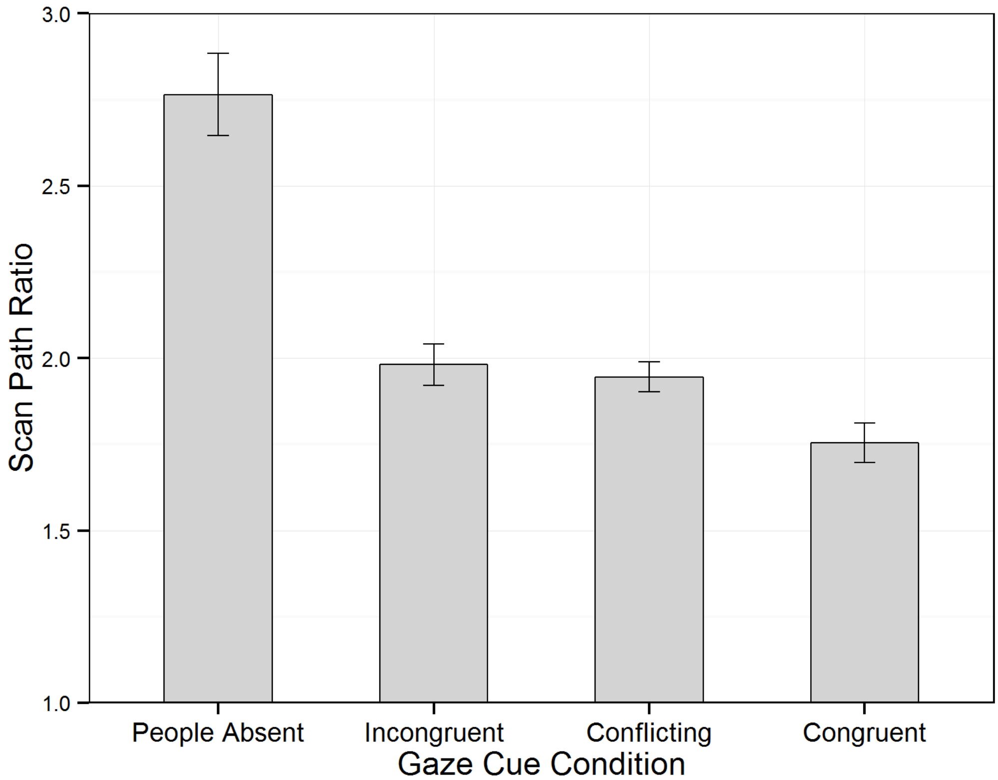

Figure 45. The scan path ratio across four gaze cue conditions. Error bars show the standard error across all data samples.

As is clear from Figure 45, the scan path ratio in the people absent condition was significantly higher than all people present conditions. After transforming the data with a logarithmic function to satisfy model assumptions, analysis showed that the people absent gaze cue condition resulted in a significantly less efficient search than the incongruent, $\beta = -0.772$, $SE = 0.138$, $t = -5.562$, conflicting, $\beta = -0.811$, $SE = 0.118$, $t = -6.883$, and congruent gaze cue conditions, $\beta = -0.996$, $SE = 0.143$, $t = -6.944$. There were also clear differences between the person present gaze cue conditions. The follow-up LMM indicates that the congruent gaze cue condition resulted in the most efficient search, producing better performance than both the incongruent, $\beta = 0.223$, $SE = 0.109$, $t = 2.053$, and conflicting gaze cue conditions, $\beta = 0.184$, $SE = 0.095$, $t = 1.920$, though the latter is only approaching significance. There were no differences between the incongruent and conflicting gaze cue conditions ($t < 0.5$).

With a definition of error as the proportion of trials in which participants made false-positive responses (that is, they pressed the button without having fixated the target object), it can be seen from Figure 46 that when people are absent from the scene false-positive responses are relatively rare, increasing in frequency once people are introduced into the scene.

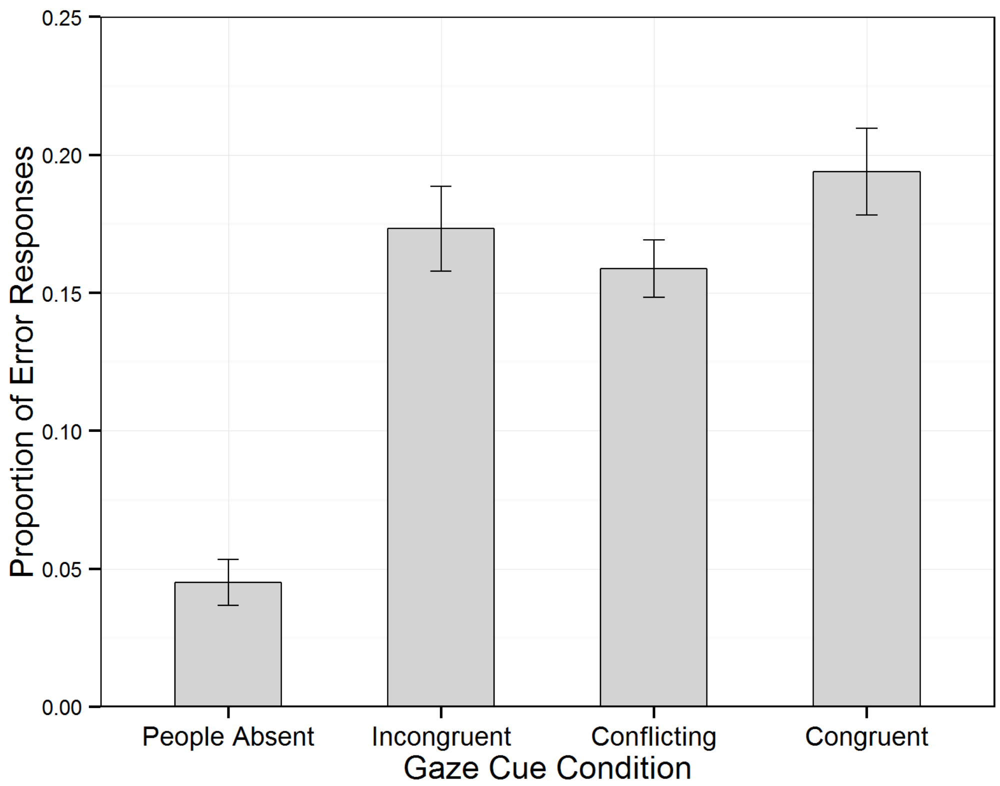

Figure 46. The proportion of false-positive responses in all trials across four gaze cue conditions. Error bars show the standard error across all data samples.

All three people present gaze cue conditions resulted in significantly more errors than the people absent condition, with the incongruent, $\beta = 0.126$, $SE = 0.023$, $t = 5.297$, conflicting, $\beta = 0.113$, $SE = 0.019$, $t = 5.858$, and the congruent gaze cue conditions, $\beta = 0.151$, $SE = 0.021$, $t = 7.170$, all producing high *t*-values. When the three people present gaze cue conditions were compared to each other, analysis showed that there was relatively little difference in the proportion of error responses between the incongruent and congruent, and incongruent and conflicting gaze cue conditions (*t*s < 1). The congruent gaze cue condition did produce a slightly higher proportion of error responses than the conflicting condition, $\beta = -0.037$, $SE = 0.024$, $t = -1.523$, but this was not significant.

Finally, the extent to which overt fixations on the person in the scene occurred was considered using ‘correct’ trials; trials where no false-positive response occurred. This is simply because in trials where participants make an error response, it is impossible to be sure of what type of strategy they were using for the search or what might have caused them to make a false-positive response. To measure the extent of overt gaze seeking, the total number of looks towards the person in the scene was calculated, with a ‘look’ defined as a fixation on the head of the person. Analysis showed that across a total of 2055 trials where a person was present and a correct response occurred, only 157 involved a fixation on the person’s face, which accounts for just 4.67% of trials.

Discussion

This study examined the effects of simultaneously presented multiple gaze cues on observer eye movement behaviour when searching for targets within real world scenes. Gaze cues were provided by two people sitting behind the table upon which the array of objects was arranged. General effects of person presence were less distinct than in studies using only one gaze cue described in earlier chapters; performance varied across gaze cue condition rather than by the presence or absence of people in the scene. These results provide a first indication of how multiple gaze cues are processed and used by observers when viewing real world social scenes.

At the initiation stage of search, measured by first saccade latency, the effects of person presence seem counter-intuitive. Here, incongruent and conflicting gaze cues resulted in faster first saccade latencies than the people absent gaze cue condition; whereas the congruent condition – which is the most helpful type of gaze cue – shows no real improvement in first saccade latency compared to when there is no gaze cue at all. With regards to the search initiation phase, this result is somewhat atypical. For the remainder of search initiation measures (first saccade direction and end point accuracy of the first saccade), congruent gaze cues resulted in the best performance of all three people present gaze cue conditions.

First saccade direction is where the greatest variation occurs between the different levels of gaze cue congruency in this stage of search. In this measure the people absent and incongruent gaze cue conditions resulted in an almost identical proportion of first saccades being directed toward the target, and these proportions were much lower than in the other conditions. As one cue is directed toward the target in the conflicting condition, a significant increase in the proportion of first saccades directed toward the target was seen. Once both cues were directed toward the target in the congruent condition, this proportion increases significantly again. This outcome is unique amongst all the measures discussed in this chapter, as it is the only one in which significant improvement was seen as more cues are directed toward the target. In other words, the degree of congruency offered by the gaze cues in the scene seems to correlate with the proportion of first saccades directed toward the target, increasing as the gaze cues become more congruent: one cue toward the target is better than none, and two cues are best of all.

In previous chapters the direction and end point accuracy of the first saccade have been used together to indicate the general accuracy of search initiation. However, it seems that once observers are presented with multiple cues, these measures no longer correlate in the same way. While the proportion of first saccades directed toward the target improved with people presence and an increase in gaze cue congruency, the end point accuracy of the same saccade – how close it brought the eyes to the centre of the target – was negatively affected by person presence. It was the people absent gaze cue condition that resulted in the most accurate landing point of the first saccade. However, the pattern of results in the people present gaze cue conditions follows the pattern seen in the first saccade direction measure. The incongruent gaze cue condition, which is the least helpful of the three people present conditions, resulted in the poorest accuracy. Accuracy gradually improved as more cues were directed toward the target, with the congruent condition resulting in accuracy more similar to that seen in the people absent condition, but which was still significantly worse than end point accuracy when there were no people in the scene.

Generally in the search initiation phase there were no significant differences between the conflicting and congruent gaze cue conditions, with benefits of congruent gaze cues emerging only when the incongruent and congruent conditions were compared. The facilitation effects of congruent gaze cues are not in question; indeed these support the findings of single-cue Posner-type tasks (e.g. Driver et al., 1999; Friesen & Kingstone, 1998; Ricciardelli et al., 2002). It is the data from the conflicting gaze cue condition that is perhaps more challenging to interpret. In the conflicting gaze cue condition participants are presented simultaneously with one cue directed toward the target, and another directed toward the distractor. While there are no effects of congruency apparent in the first saccade latency measure, in first saccade direction and the end point accuracy of the first saccade the conflicting condition lies numerically between the incongruent and congruent conditions, yet it is only statistically different from the incongruent condition. This would suggest that the first saccade of search benefits as much from two congruent cues as from one congruent cue and one incongruent cue.

There are two possible explanations for this result. First, it may be that the similarity between measures of search initiation in the conflicting and congruent gaze cue conditions suggests participants did not process and utilise multiple cues simultaneously in order to programme their first saccade. Had both cues been processed simultaneously, more accurate first saccades would be expected when participants were presented with two helpful congruent cues than when they were given one helpful and one unhelpful cue in the conflicting gaze cue condition. However, this explanation does not adequately account for the differences seen in the first saccade direction measure. Here, the congruent cue resulted in improved accuracy when compared to the conflicting condition in terms of the probability of the first saccade being directed toward the target. This result suggests both cues were used and processed simultaneously when programming the first saccade, as otherwise this statistically significant difference would not occur. Ultimately, comparing the conflicting condition to either the incongruent or congruent gaze cue condition cannot definitively determine how participants are using the simultaneously presented gaze cues.

An alternative interpretation of these results could suggest that the first saccade is programmed with respect to gaze cue information, but also with respect to matching visual characteristics of the scene to information held within the internal search or target template (e.g. Rao, Zelinsky, Hayhoe & Ballard, 2002). Malcolm and Henderson (2009) define the target template as an individual's internal representation of a target object. This template contains prominent features of an object, for example the representation of a mug may be a cylindrical shape with a prominent handle feature, and this type of template becomes increasingly important when searching for a target object amidst a cluttered scene with multiple distractor items. More detailed knowledge of the target features results in greater facilitation of search (Castelhano & Heaven, 2010; Malcolm & Henderson, 2009), and these target features weight spatial locations in the scene according to the degree to which they match with the template, helping to guide subsequent eye movements in the visual search (Wolfe, Horowitz, Kenner, Hyle & Vasan, 2004). Spotorno et al. (2014) investigated how the target template could influence eye movement behaviour in this first epoch of search, containing programming of the first saccade, by presenting participants with scenes in which the target object appeared in either highly-probably or highly-improbable locations (e.g. a cow appeared in a field, or in the sky). When the target object appeared in a highly-probable location, participants' performance was facilitated with a higher proportion of first saccades direction toward the target. These results show that the target template can facilitate search even in programming the first saccade, and that search initiation involves parallel processing of visual information and matching to the internal template prior to launching the first saccade. Considering this in terms of the current study, it is possible the initial weighting of the scene occurs from both processing of visual characteristics of the scene, but also from processing of the gaze cues provided by the people in the scenes. Simply having people present in the scene influences search behaviour, even at the very first saccade. It is possible to suggest that the gaze cues provided in people present scenes are additional cues weighting attention to multiple locations within the scene (Awh & Pashler, 2000; Casteillo & Umiltà, 1990).

If gaze cues are acting as additional means of weighting locations within the scene for subsequent search, attention must be divided appropriately in order to track multiple objects simultaneously (Pylyshyn & Storm, 1998; Scholl, Pylyshyn & Feldman, 2001), to compare different locations (Awh & Pashler, 2000), or to process distinct areas of the scene simultaneously (Castiello & Umiltà, 1992). Attention has been described as a 'spotlight'; much like the spotlight on a stage highlighting a performer, the spotlight of our attention moves to focus on the stimulus to be processed – anything falling under the beam of the spotlight would be processed more efficiently (Posner, 1980; Umiltà, 1988). Alternatively, we might consider attention to operate like the zoom lens of a camera (Eriksen & St. James, 1986; Eriksen & Yeh, 1985), which can 'zoom in' to focus on a single stimulus, or 'zoom out' to include several nearby stimuli within its range. What both of these theories have in common is that they consider the spotlight or zoom lens of attention to be singular – how then do they account for parallel processing of multiple cues? Castiello and Umiltà (1992) would suggest that in the current study, the spotlight of attention is distributed over a wide enough area that can encompass both spatial locations cued by the people present in the scene. Shulman et al. (1979) would instead propose that the spotlight of attention alternates rapidly between the two cued spatial locations. The problem with a singular focus of attention is that as the area to which it is distributed is broadened, the detail to which the stimuli can be processed deteriorates. The solution to this is, of course, to have more than one spotlight of attention.

McMains and Somers (2004) investigated the possibility of multiple spotlights of attention that could be used to process two or more separate and distinct spatial locations simultaneously. They cite the longstanding debate regarding how attention is allocated, and the difficulty of determining attention allocation using only behavioural methods. To that end, McMains and Somers (2004) used functional Magnetic Resonance Imaging (fMRI) to monitor activity in the visual cortex while participants performed a visual monitoring task that required them to monitor two spatially distinct areas simultaneously. Their results highlighted the distribution of attention in response to task demands, revealing two distinct spotlights of attention operating in parallel. In a follow up study, again using fMRI to monitor neural activity, McMains and Somers (2005) determined that deploying attention via these multiple spotlights was surprisingly efficient, and that spatial attention can be easily deployed across a wide range of configurations.

Considering the pattern of results in the search initiation phase, it is possible to conclude that the current study suggests greater weighting resulting from visual features (detected by the target template) than spatial gaze cues. If this is indeed the case, a single gaze cue toward a distractor – as in the conflicting gaze cue condition – may not impact to any great extent on saccadic programming, but two coincident cues directed toward the distractor – in the incongruent condition – may add sufficient weighting to this distractor location to interfere with the programming of the first saccade. This would explain why in the search initiation phase only the incongruent gaze cue condition resulted in less accurate initial saccades.

The findings of scene scanning stage of search are broadly consistent with this proposed framework. However, in this second stage of search, the conflicting gaze cue condition performs more similarly to the incongruent condition where both cues are unhelpful, than to the congruent condition where two helpful cues are provided. This is particularly evident in measures exploring first detection of the target (time taken to first fixation on the target) and search efficiency (scan path ratio). In these measures the conflicting condition was statistically identical to the incongruent condition, but much greater differences are seen between the conflicting and congruent conditions. As performance in the conflicting gaze cue condition seems to deteriorate in the scene scanning phase, it could be suggested that incongruent cues have a much greater effect in the later stages of search. Indeed, a single incongruent cue – as is seen in the conflicting gaze cue condition – is sufficient to significantly increase the time taken until the first fixation on the target compared to when both cues are directed toward the target in the congruent gaze cue condition.

To proceed with the assumption that multiple gaze cues are processed simultaneously, and that they lead to addition weighting of stimuli in gazed-at locations, it must be the case that gaze cuing effects in the present study are not dependent on overtly selecting a gaze cue with foveal vision; instead gaze cues can be detected and responded to covertly (Knoeferle & Kreysa, 2012; Macdonald & Tatler, 2013a). This is reflected by the data concerning fixations on either person's face within the scene. Participants' gaze was very rarely directed toward either individual depicted in the scene, with only 4.38% of people present trials featuring at least one fixation on either person's face in all three people present gaze cue conditions. In the current study, there was no reason indicated by the instructions given to participants that participants should look at the people in the scenes to complete the task. Assuming gaze can be followed covertly, fixating on the table in order to find the target quickly (as speed was emphasised by task instruction) whilst covertly accessing gaze information is an appropriate response.

The present study provides strong evidence that multiple simultaneous gaze cues influence visual search in a number of ways. In the initial stages of search, more congruent – and therefore more helpful – gaze cues resulted in more accurate first saccades that were both more likely to be directed toward the target, and brought the eyes closer to the centre of the target. In the second stage of search, incongruent cues had a much stronger effect on eye movement behaviour. Evidently even one incongruent cue was enough to disrupt search, with the conflicting gaze cue condition resulting in similar levels of performance as the incongruent condition. It has been suggested that when participants are presented simultaneously with multiple gaze cues in the visual search task, they use two processes to allocate weighting to separate locations within the scene. Gaze cues are processed and utilised in parallel even prior to search initiation, however they contribute a lesser weighting to guide subsequent search than the visual features of stimuli matching to the internal target template. Consistent with this framework, the facial region of individuals depicted in the scenes were rarely fixated, suggesting the effects of gaze cues were the result of covert peripheral processing of these cues. This is the first study to examine the processing of multiple gaze cues within a social scene and offers first insight into how we might process multiple cues when they are presented simultaneously. However, to properly recreate Posner-type paradigms, instruction must be given with reference to the purpose of having people present (or absent) within the search scenes. As in the development of the single cue paradigm, the logical next step in using this multiple cue paradigm would be to explore how instructions regarding person presence affect visual search behaviour. This will be explored in the next chapter.

Chapter Six

How do manipulations of perceived helpfulness of simultaneously-presented multiple gaze cues impact participant eye movement behaviour during a visual search task?

Introduction

The previous chapter expanded on the newly developed realistic Posner-type paradigm by adding simultaneously-presented multiple gaze cues. The need to create stimuli that have greater ecological validity is becoming increasingly important (e.g. Risko et al., 2012), which is why it is necessary to present participants with scenes that more closely resemble what they would experience in the real world, particularly if we wish to ascribe the results obtained to real world behaviour. In the same way as the original single-cue paradigm was expanded upon, this chapter explores how two different types of instruction affect participants' eye movement behaviours during a multiple-cue paradigm.

Previous Posner-type studies that found evidence of reflexive gaze following (e.g. Driver et al., 1999; Friesen & Kingstone, 1998; Ricciardelli et al., 2002) suggest a bottom-up process of gaze following; it is automatic and beyond our control. This is evidenced by instructions given to participants in these studies that inform them the gaze cues provided are not helpful for the task at hand, yet gaze following still occurs. It might be argued then that were gaze following a top-down process, there would be an inhibition of gaze following after participants are made aware the cues are not helpful. This model of search being a stimulus-driven process (bottom-up) has strong support. There is a well-established foundation of literature that shows visual preference for contrast and edges (Henderson & Hollingsworth, 1999), ‘pop-out’ – the effect of one item being visually distinct from its surrounding distractors (e.g. Treisman & Gellade, 1980; Wolfe, Butcher, Lee & Hyle, 2003). However, as has been discussed in previous chapters, once more realistic methodologies are adopted evidence for this clearly defined reflexive response begins to lessen, and indications of reflexive orienting are not as clear cut.

When this is further expanded by adding a top-down manipulation to the task in the form of a helpful or unhelpful instruction, hypotheses about the anticipated performance of participants in the task become more complex. While research that investigates task instructions (e.g. Schwartz et al., 2005; Yarbus, 1967) shows clear effects of the instruction given on the participants’ performance, the addition of instruction to the novel paradigm created for this thesis – as discussed in Chapters Three and Four – showed no effects of instruction. Participants’ eye movements and overall performance were consistent across the no instruction, helpful instruction, and unhelpful instruction conditions. This would suggest that when only one person provides a gaze cue in a realistic scene, task instructions do not impact on observer eye movement behaviour. A study conducted by Greene et al. (2009) using fMRI to monitor brain activity whilst participants completed a spatial cuing task suggests that the automatic effects of gaze occur regardless of instructions given to participants. The neurological evidence gathered in their study indicates that the special status of gaze, and the isolated processing of gaze stimuli, is irrefutable and cannot be overridden by top-down processes.

However, the evidence supporting effects of task instruction cannot be completely disregarded. There is consistent evidence across a number of studies that demonstrate how instructions given to participants can manipulate their performance in the task. In particular, the study conducted by Itier et al. (2007) is arguably more methodologically similar to the current paradigm than the fMRI study conducted by Greene et al. (2009). Itier et al. (2007) used a Posner-type presentation where participants were shown faces with either direct or averted gaze, and with heads presented either facing straight forward or in a three-quarter view and asked to detect either head or gaze direction. There was evidence of gaze processing in both tasks, suggesting there is some agreement with Greene et al.'s (2009) findings that gaze will be processed in all tasks, but the extent to which gaze was fixated or followed varied depending on the task given to participants. Approximately 90% of participants' first saccades landed in the eye region of the face during the gaze location task compared to less than half in the head orientation task. Itier et al. (2007) argued that based on their findings, participants responses cannot be directed solely by a reflexive gaze-orienting mechanism and that instead there must be some extent to which exogenous variables can impact on search behaviour.

It is possible that the nature of the task itself causes some variation in the degree to which instructions may affect participant behaviour. More simple Posner-type tasks may elicit different eye movement behaviour than more complex real world studies. While the single cue paradigm with varied instruction explored in Chapter Three did not produce any effects of instruction, it has already been seen in Chapter Five that providing participants with multiple simultaneously-presented gaze cues evokes a different response than a single cue in an identical environment. If participants use different search strategies when two cues are present as opposed to one, it is quite possible that instruction will have different effects when two cues are present than when participants are given a single cue. This chapter explores how helpful and unhelpful instructions given prior to commencing the search task impact on participants' search. As in previous studies, one condition instructs participants to ignore the presence of a person in the scene (the unhelpful instruction). The second condition addresses an oversight in the previous literature, by examining what happens when participants are told the helpful reason for person presence: that they may be helpful in completing the task as they may be looking at the target. This helpful instruction manipulation will be the first to explore a helpful instruction given to participants when more than one gaze cue is present.

Method

Participants

A total of 20 people (9 male) were recruited for the unhelpful instruction condition, and a further 20 people (7 male) were recruited for the helpful instruction condition in these studies. All had normal or corrected vision and were naïve to the purposes of the study. Level one and two undergraduate students received course credits for participation; anyone not eligible for course credit was paid £2. Individuals recruited for participation in these studies had not participated in any of the previous experiments

Materials

The materials used were the same as discussed in Chapter Five. However, to give a quick review, experimental scenes were created using ten different sets of everyday objects. Each scene featured one of the ten sets of 15 everyday items arranged on a table top. Within each scene one item was designated the target and another designated the distracter. These items were always on the opposite sides of scene centre, preventing any central bias (as discussed in Tatler, 2007). The target was equally likely to appear on the left or right side of the table. Every arrangement was photographed four times, to create a total of three people present gaze cue conditions: the *congruent* gaze cue condition, where both people cue the target; the *incongruent* condition, where both people cue the distractor; and the *conflicting* gaze cue condition, where one person cues the target and the other cues the distractor. Two versions of the conflicting gaze cue condition were created such that the individual cueing the target could be counterbalanced within and across participants. In addition to the four person present photographs for each arrangement, a final arrangement was photographed without any people present to create a people absent control scene. In total ten people absent scenes and 160 person present scenes were created.

Eye Tracking

Eye movements were tracked using the same equipment as stated in previous chapters. The same procedure was followed to accurately calibrate the desk-mounted eye tracker, and the same acceptance criteria (average spatial error less than 0.5 degrees and maximum error less than 1 degree over the 9 calibration points) were used.

Procedure

Trials followed the same procedure as in previous chapters. To recap: a single-point calibration check was performed before each trial began. The name of the target object was presented for 500 ms, followed by a blank screen for 500 ms, and then the presentation of the scene for a maximum of 10 s, or until the trigger button on the gamepad was pressed by the participant indicating they had found the target object. Each participant saw a total of 100 scenes: 20 people absent scenes and 80 person present scenes. This means that every participant saw every image twice, once searching for the target object and once searching for the distractor. The difference between versions one and two of the experiment was simply which of the two photos within an arrangement (as in Figure 39) were presented.

The instructions given to participants, beyond the basics of how to perform the task, were designed to manipulate the perceived helpfulness of the gaze cues provided in people present scenes. In the unhelpful instruction condition, participants were told to ignore the presence or absence of a person in the scene, similar to previous Posner-type tasks, with the instruction: "*Some of the scenes will have people in them, but please just ignore them. I'm using the same images over several experiments, but in this experiment the people are not relevant; I'm only interested in how you search for the target object in the scene.*" Conversely, in the helpful instruction condition, participants were told: "*Some of the scenes will have people in them. These people might be looking at the target, so they may help you find it faster.*" This instruction does not tell participants that they must look at the people; it simply provides them with more information about the context of the scene.

Data Analysis

Following the structure of analyses in previous chapters, data were analysed using linear mixed effects models (LMMs) in the lme4 package (Bates et al., 2011) within the R statistical analysis environment (R Development Core Team, 2011). In linear models, the lmer() function returns t-values without the associated p-values, so any effects for which the t-value is greater than two – that is effects larger than twice their standard error – as reflecting a significant effect (as in Kleigl et al., 2012). In the analyses below, it is noted if any data required logarithmic transformation to generate a normal distribution, and the model with the most complicated random effects structure that converged was reported. As in Chapter Five measures of search initiation (first saccade latency, direction, and end point accuracy) consider data from all trials, irrespective of whether the participant fixated the target, as did the measure of error rate. The time to first fixation, response time, scan path ratio, and overt gaze-seeking analyses used only correct-response trials, where participants had both fixated the target and pressed the trigger button to indicate they had found it. It should be noted that in the results section below the two instruction conditions will be considered separately; the analyses consider effects of gaze cue type within each instruction type.

Results

Unhelpful Instruction

Search Initiation

The first measure of search initiation to be considered is the time taken to launch the first saccade after the appearance of the scene (first saccade latency). For analysis first saccade latency required two transformations. The data presented a small number of very short latencies, which most likely were the results of pre-emptive eye movements beginning before the appearance of the scene. Very short latencies were defined as being less than 100 ms. A total of 296 trials featured a very short latency (14.8% of the total number of trials), and these very short latencies were removed. The remaining data underwent logarithmic transformation to generate a normal distribution. The first LMM compared the three people present gaze cue conditions to the people absent condition. As can be seen in Figure 47, there was very little different between these conditions with only the incongruent gaze cue condition demonstrating a different first saccade latency than the people absent condition, $\beta = -0.016$, $SE = 0.009$, $t = -1.74$, but this did not reach significance. Neither the conflicting nor congruent condition showed any real difference from the people absent condition (ts < 1).

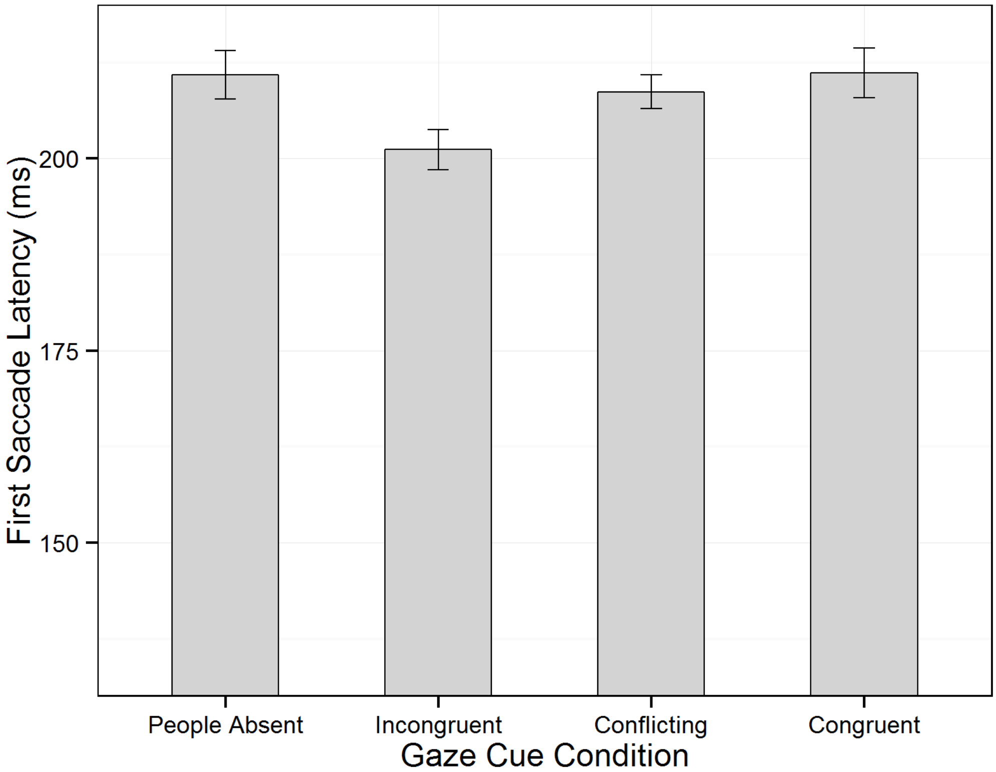

Figure 47. First saccade latency in each of the four gaze cue conditions. Error bars show standard error across all data samples.

When comparing the people present gaze cue conditions to each other, the incongruent gaze cue condition clearly resulted in the shortest first saccade latencies. This condition produced shorter first saccade latencies than both the conflicting, β = 0.011, *SE* = 0.006, *t* = 1.71, and congruent gaze cue conditions, β = 0.014, *SE* = 0.007, *t* = 1.90, though neither of these reached significance. There were no differences in first saccade latency between the conflicting and congruent gaze cue conditions ($t < 0.5$).

The proportion of first saccades directed toward the target showed more variation across the four gaze cue conditions, as can be seen in Figure 48. First saccades directed toward the target were in similar proportions in the people absent and incongruent conditions ($z < 1$, $p > 0.5$). While there was numerically a small increase in the proportion of first saccades directed toward the target in the conflicting gaze cue condition as opposed to the incongruent condition, this was not significant ($z < 1$, $p > 0.1$). It was only once the people absent and congruent gaze cue conditions were compared that a significant benefit of person presence emerged, with the congruent cue resulting in a higher proportion of first saccades directed toward the target, $\beta = 0.366$, $SE = 0.175$, $z = 2.085$, $p < 0.05$, than the people absent condition.

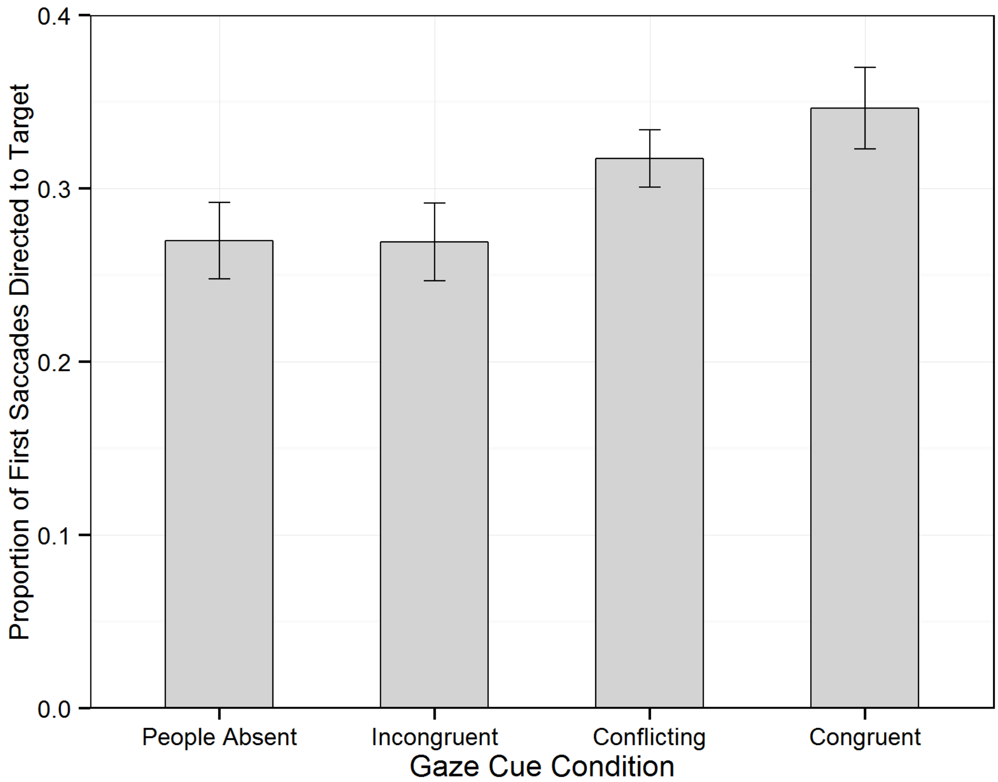

Figure 48. The proportion of first saccades directed toward the target across four gaze cue conditions. Error bars show standard error across all data samples.

Comparing the three people present gaze cue conditions showed similar differences, given that the incongruent condition was remarkably similar to the people absent condition. When compared to the incongruent gaze cue condition both the conflicting, $\beta = 0.034$, $SE = 0.028$, $t = 1.235$, and congruent gaze cue conditions, $\beta = 0.071$, $SE = 0.032$, $t = 2.241$, showed an increased proportion of first saccades directed toward the target, however only the latter reached significance. There were no significant differences in the proportion of first saccades directed toward the target between the conflicting and congruent conditions ($p > 0.1$).

The final measure of search initiation, the end point accuracy of the first saccade, showed the greatest variation between gaze cue conditions, which can be seen in Figure 49 below. This measure indicates how close the first saccade brought the eyes to the centre of the target, and as is evident from Figure 49, incongruent cues significantly interfered with first saccade accuracy. When compared to the people absent condition the incongruent gaze cue condition, $\beta = 1.696$, $SE = 0.394$, $t = 4.301$, conflicting condition, $\beta = 1.075$, $SE = 0.339$, $t = 3.163$, and congruent condition, $\beta = 0.674$, $SE = 0.389$, $t = 1.732$, were all less accurate, though in the case of the congruent condition this difference was not significant.

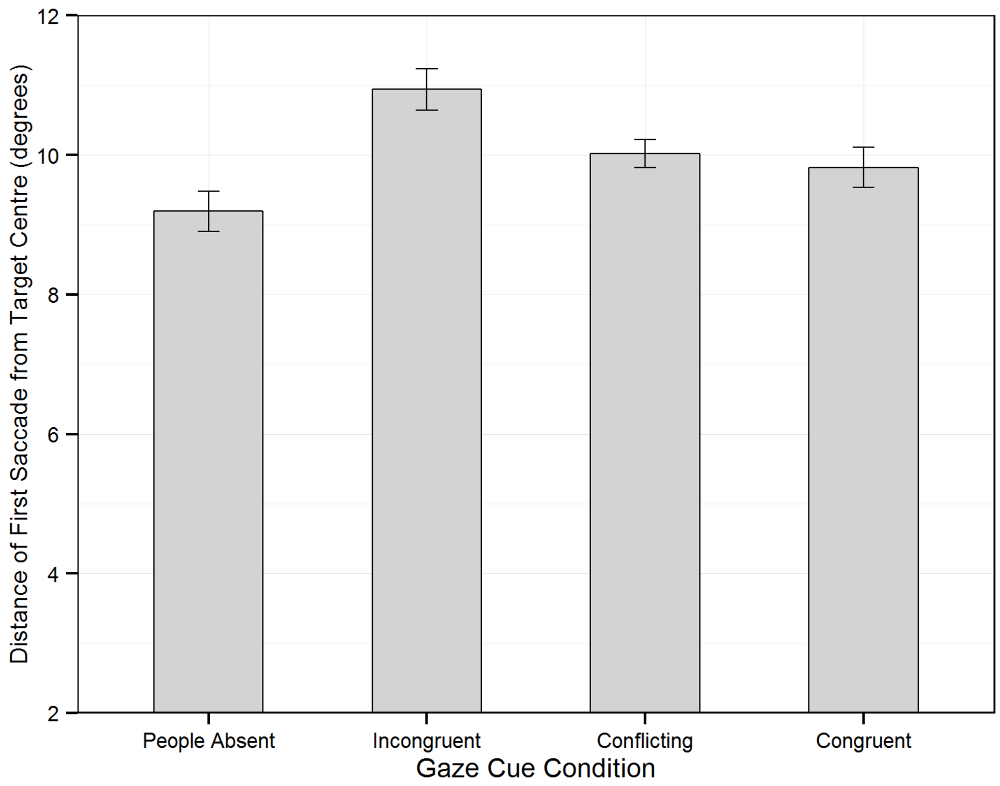

Figure 49. The distance of the landing point of the first saccade from the centre of the target ROI (in degrees of visual angle) as a measure of end point accuracy across three gaze cue conditions. Error bars show standard error across all data samples.

The incongruent gaze cue condition was the least accurate of the three, which was reflected in the follow up analyses. The incongruent condition was less accurate than both the conflicting, $\beta = -0.670$, $SE = 0.535$, $t = -1.253$, and congruent gaze cue conditions, $\beta = -1.095$, $SE = 0.420$, $t = -2.604$, though this decrease in accuracy was only significant when compared to the congruent gaze cue condition. The conflicting and congruent conditions produced similar levels of end point accuracy ($t < 1$).

Scene Scanning

Person presence was a clear benefit to performance in the measure examining time taken for participants to first fixate on the target. As can be seen in Figure 50, once people are present within the scene, this first fixation occurred much sooner. To satisfy model assumptions, logarithmic transformation of the data was required. When compared to the people absent gaze cue condition, the incongruent condition resulted in very similar times to first fixation on the target ($t < 0.5$), but the conflicting, $\beta = -0.057$, $SE = 0.025$, $t = -2.29$, and congruent gaze cue conditions, $\beta = -0.062$, $SE = 0.029$, $t = -2.10$, both resulted in significantly earlier first fixations on the target object.

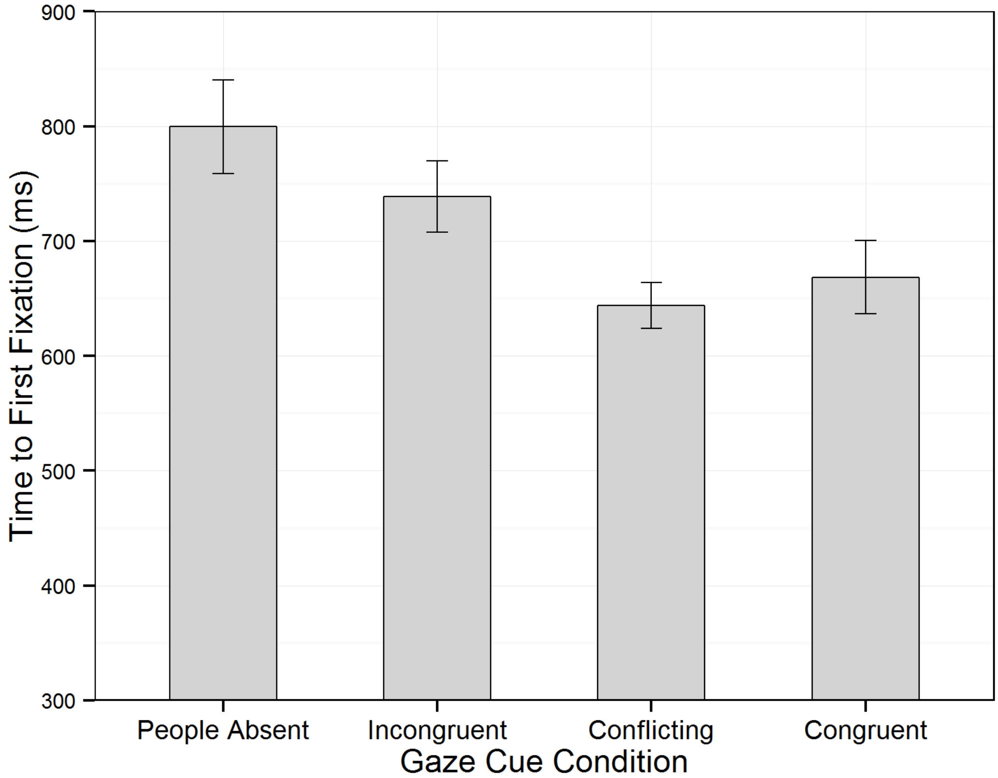

Figure 50. The time to first fixation on the target (ms) from scene presentation across four gaze cue conditions. Error bars show standard error across all data samples.

In follow up LMMs comparing the three people present gaze cue conditions, it was found that both the conflicting, $\beta = -0.053$, $SE = 0.021$, $t = -2.45$, and congruent gaze cue conditions, $\beta = -0.054$, $SE = 0.024$, $t = -2.19$, produced earlier first fixations on the target than the incongruent gaze cue condition. When the conflicting and congruent conditions were compared no difference was found ($t < 0.05$).

After logarithmic transformation to satisfy model assumptions, the first LMM was run on the response time data to compare the people absent and people present gaze cue conditions. There were clear effects of people presence with significant reductions in response time in these conditions compared to the people absent condition. The incongruent, $\beta = -0.105$, $SE = 0.021$, $t = -4.91$, conflicting, $\beta = -0.123$, $SE = 0.019$, $t = -6.38$, and congruent gaze cue conditions, $\beta = -0.127$, $SE =$

0.021, t = -5.89, all produced significantly shorter response times than the people absent condition. These results are reflected in Figure 51 below.

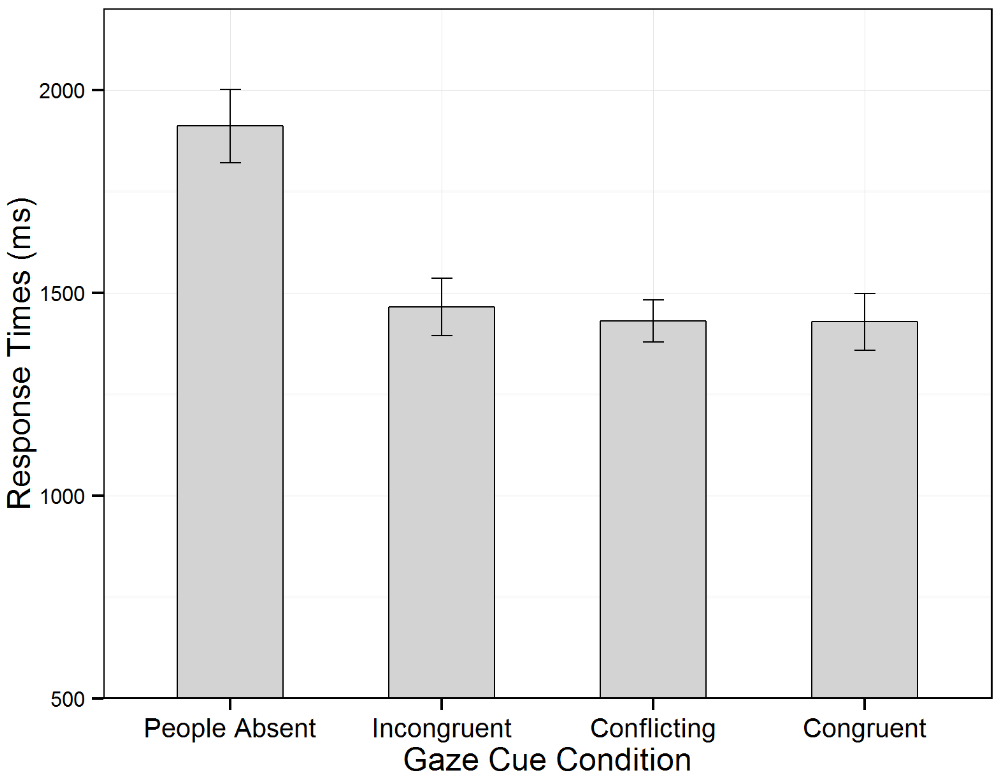

Figure 51. Response times (ms) to button press indicating successful search for the target across four gaze cue conditions. Error bars show standard error across all data samples.

When comparing across people present conditions in the follow up analysis, the differences in response times produced by each condition were much smaller. The incongruent gaze cue condition resulted in slightly longer response times than the conflicting, β = -0.018, SE = 0.018, t = -1.01, and congruent gaze cue conditions, β = -0.021, SE = 0.020, t = -1.09, but these differences were not significant. There was an even smaller difference between the conflicting and congruent gaze cue conditions (t < 0.5).

Considering scan path ratio, there were again clear benefits of people presence. Compared to the people absent gaze cue condition, the incongruent, β = -0.877, SE =

0.216, t = -4.060, conflicting, β = -0.989, SE = 0.186, t = -5.317, and congruent gaze cue conditions, β = -1.011, SE = 0.201, t = -5.014, all resulted in more efficient searches with scan path ratios closer to 1. This is demonstrated below in Figure 52.

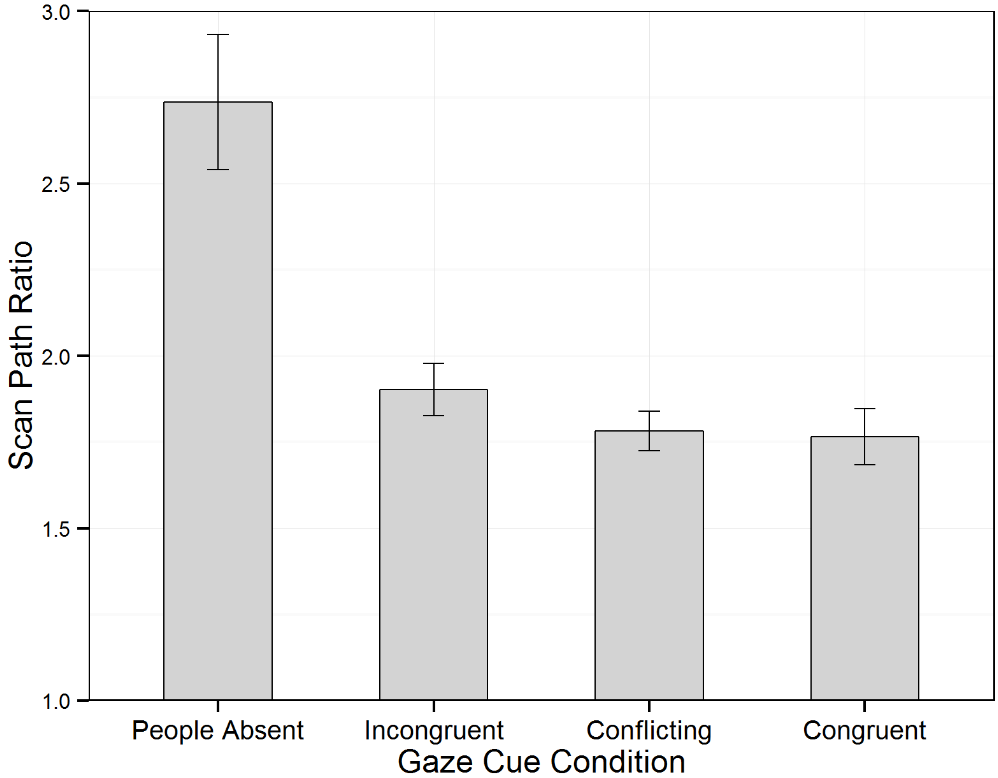

Figure 52. The scan path ratio across four gaze cue conditions. Error bars show the standard error across all data samples.

As one might expect based on the data presented in Figure 52 there was very little difference in search efficiency between the people present gaze cue conditions. When compared it was found that there were no differences between any of the conditions (ts < 1).

Error rate is the final measure for statistical analysis in the scene scanning stage of search. The proportion of false-positive responses is shown below in Figure 53.

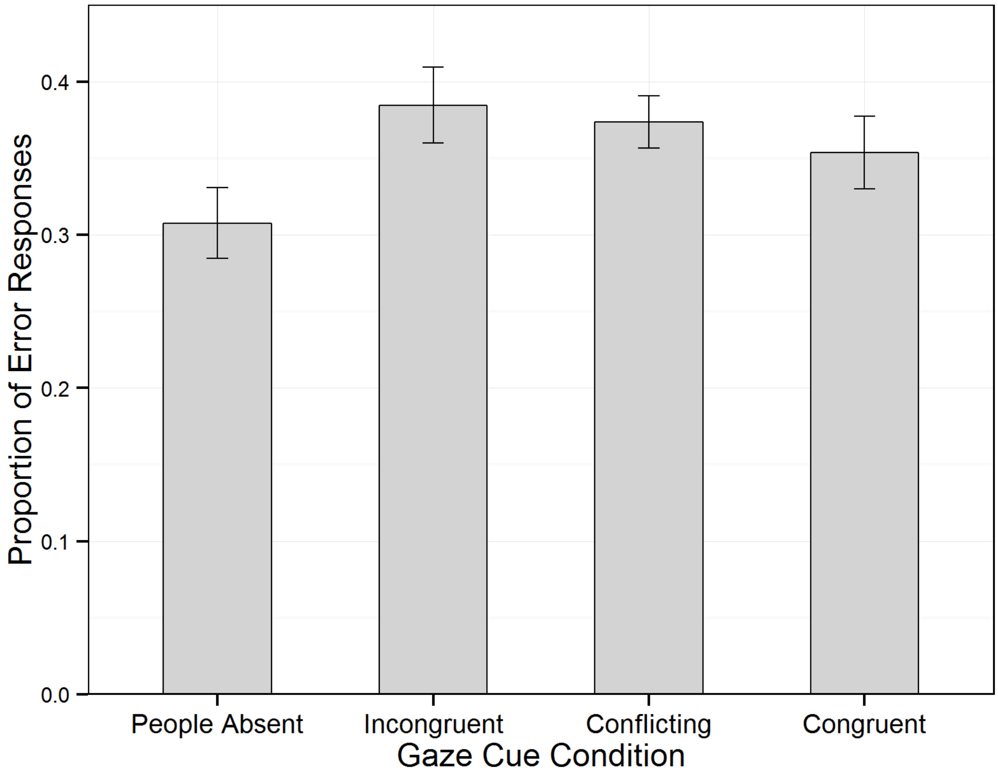

Figure 53. The proportion of false-positive responses in all trials across four gaze cue conditions. Error bars show the standard error across all data samples.

As can be seen in Figure 53, the people absent gaze cue condition actually results in the smallest proportion of false-positive responses. Both the incongruent, $\beta = 0.075$, $SE = 0.034$, $t = 2.216$, and conflicting gaze cue conditions, $\beta = 0.071$, $SE = 0.029$, $t = 2.447$, resulted in a significantly higher proportion of errors than the people absent condition. While the congruent gaze cue condition still had a higher proportion of errors than the people absent condition, $\beta = 0.047$, $SE = 0.034$, $t = 1.360$, this was not significant. The follow up analyses showed there were no differences in the proportion of errors across the three people present gaze cue conditions (ts < 1).

Finally, the extent to which overt fixations on the person in the scene occurred were considered using ‘correct’ trials; trials where no false-positive response occurred. This is simply because in trials where participants make an error response, it is impossible to be sure of what type of strategy they were using for the search or what might have caused them to make a false-positive response. To measure the extent of overt gaze seeking, the total number of looks towards the person in the scene was calculated, with a ‘look’ defined as a fixation on the head of the person. Analysis showed that across a total of 1006 trials where a person was present and a correct response occurred only 21 involved a fixation on the person’s face, which accounts for just 2.08% of trials.

Helpful Instruction

Search Initiation

First saccade latency analysis required two transformations. The data presented a small number of very short latencies, which most likely were the results of pre-emptive eye movements beginning before the appearance of the scene. Very short latencies were defined as being less than 100 ms. A total of 296 trials featured a very short latency (14.807% of the total number of trials), and these very short latencies were removed. The remaining data underwent logarithmic transformation to generate a normal distribution. These data are presented in their original form in Figure 54.

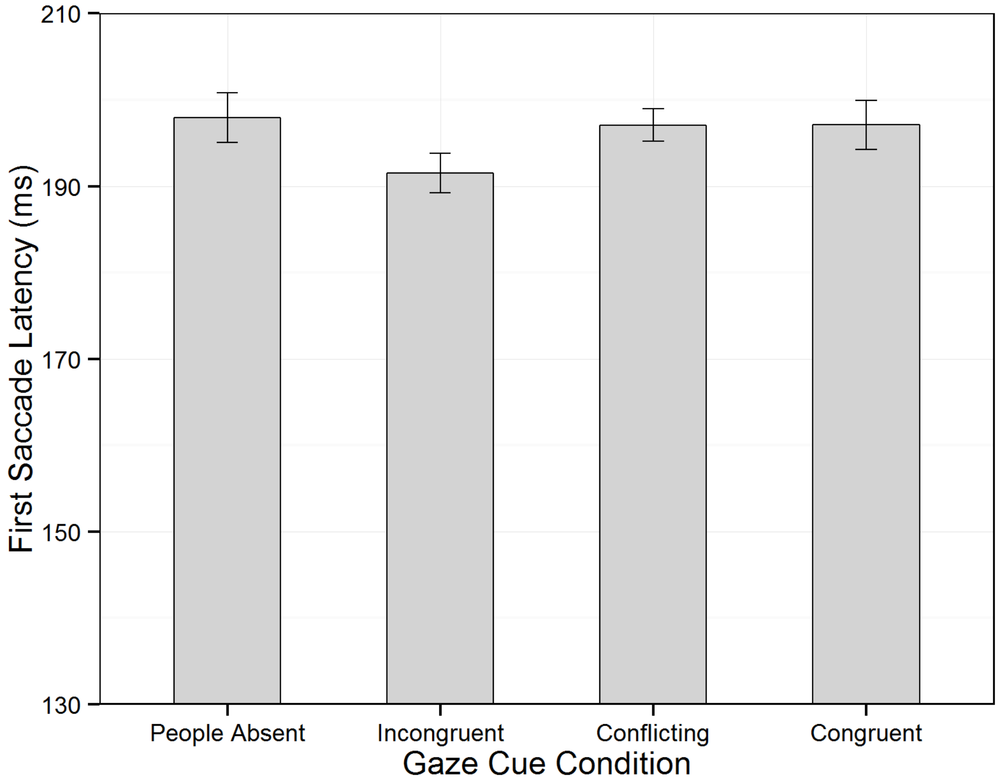

Figure 54. First saccade latency in each of the four gaze cue conditions. Error bars show standard error across all data samples.

When compared to the people absent gaze cue condition, only the incongruent gaze cue condition showed any real difference in the length of the first saccade latency, β = -0.010, *SE* = 0.008, *t* = -1.25, with a slightly shorter first latency, but not a great enough difference to produce a significant effect. The first saccade latencies produced by the conflicting and congruent conditions were almost identical to the people absent gaze cue condition (*t*s < 0.05). As the incongruent gaze cue condition produced the shortest first saccade latency, there were differences apparent when this condition was compared to both the conflicting, β = 0.010, *SE* = 0.006, *t* = 1.53, and congruent gaze cue conditions, β = 0.009, *SE* = 0.007, *t* = 1.36, though neither of these differences were significant. First saccade latencies in the conflicting and congruent gaze cue conditions were almost identical (*t* < 0.05).

The proportion of first saccades directed toward the target showed a significant increase once people were present in the scene, as can be seen from Figure 55. When compared to the people absent gaze cue condition there was some increase in the proportion of first saccades directed toward the target apparent in the incongruent condition, $\beta = 0.335$, $SE = 0.215$, $z = 1.561$, $p = 0.118$, although this was not significant. However, once at least once congruent cue is provided the effect becomes stronger with both the conflicting, $\beta = 0.390$, $SE = 0.178$, $z = 2.189$, $p < 0.05$, and congruent gaze cue conditions, $\beta = 0.478$, $SE = 0.188$, $z = 2.545$, $p < 0.05$, resulting in significantly more first saccades directed toward the target than the people absent gaze cue condition.

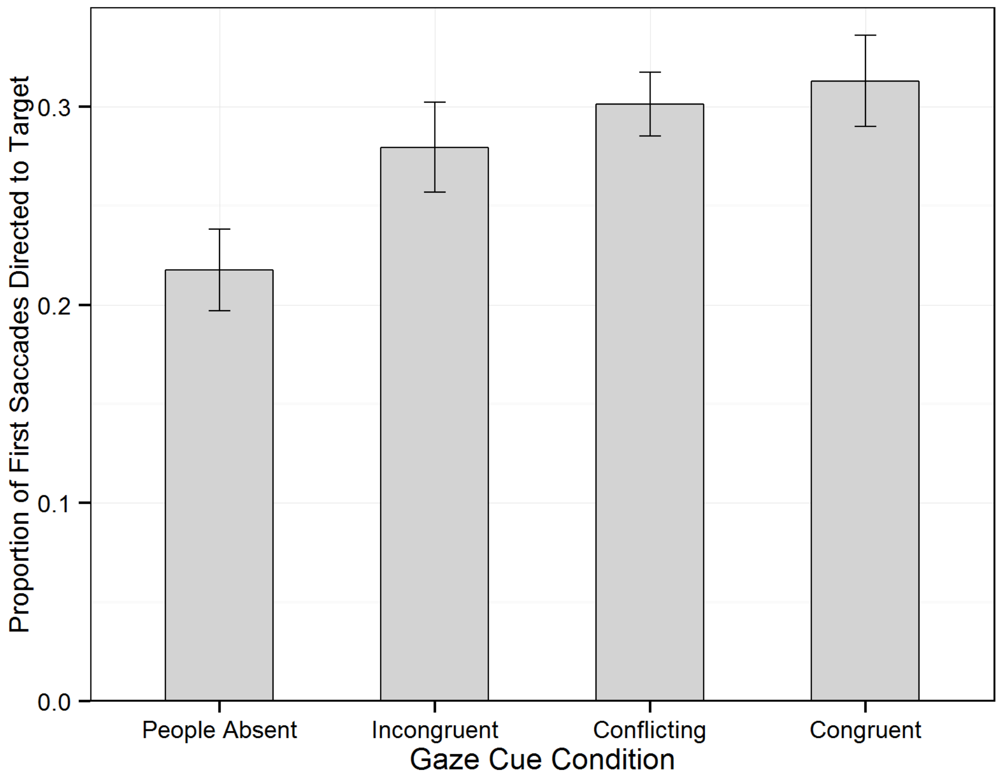

Figure 55. The proportion of first saccades directed toward the target across four gaze cue conditions. Error bars show standard error across all data samples.

Follow up analyses comparing first saccade direction across the three people present gaze cue conditions showed that there were no significant differences between these groups (*t*s < 1).

The end point accuracy of the first saccade, which is the final measure of the search initiation phase, provides an indication of how close the first saccade brought the eyes to the target. When compared to the people absent condition, the incongruent, $\beta = 1.202$, $SE = 0.379$, $t = 3.170$, and conflicting gaze cue conditions, $\beta = 0.742$, $SE = 0.326$, $t = 2.273$, both resulted in significantly less accurate first saccades than the people absent condition. The congruent condition resulted in end point accuracy that was closer to the level of the people absent condition, $\beta = 0.617$, $SE = 0.374$, $t = 1.647$, which while not significant, still shows quite a considerable difference. These data are presented in Figure 56.

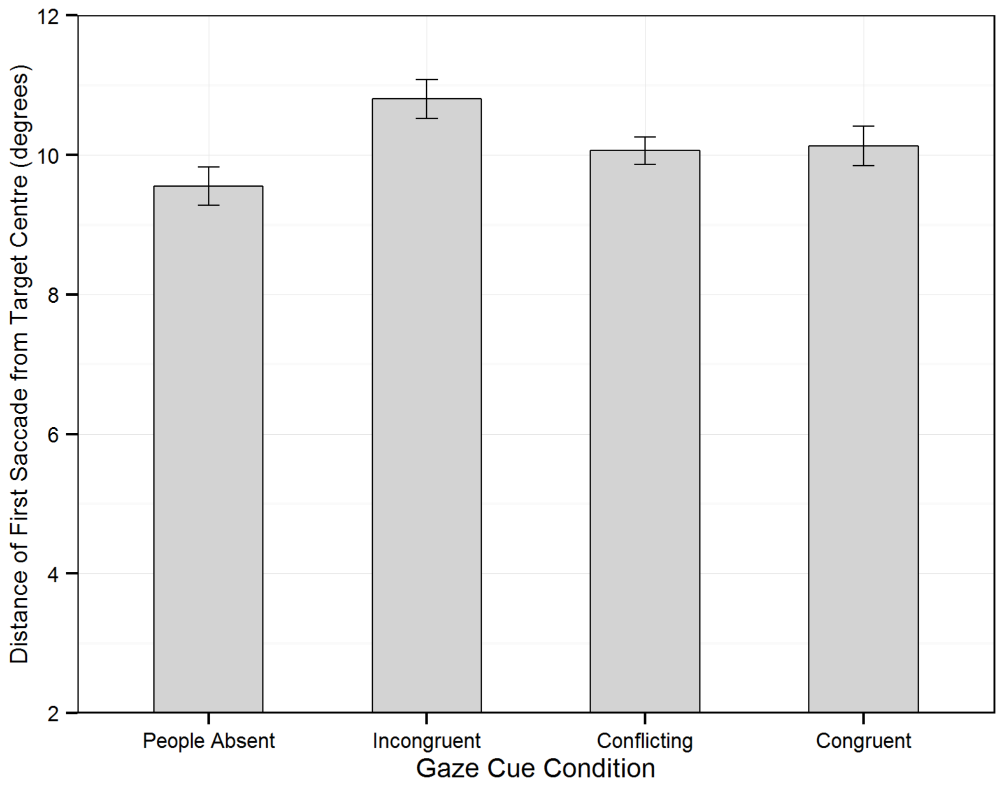

Figure 56. The distance of the landing point of the first saccade from the centre of the target ROI (in degrees of visual angle) as a measure of end point accuracy across four gaze cue conditions. Error bars show standard error across all data samples.

Further analyses showed that there was some improvement in end point accuracy in the congruent condition as compared to the incongruent condition, β = 0.663, *SE* = 0.422, *t* = 1.571, and the conflicting condition as compared to the incongruent condition, β = -0.475, *SE* = 0.424, *t* = -1.119, though neither of these results were significant. There were no differences between the conflicting and congruent gaze cue conditions ($t < 1$).

Scene Scanning

The time taken until the first fixation on the target object provides an indication of how long it took participants to locate the target, even if they do not press the trigger button on the gamepad at this point. After logarithmic transformation to satisfy model assumptions, the analysis showed that people presence significantly benefitted performance in this area. When compared to the people absent gaze cue condition the incongruent, β = -0.063, *SE* = 0.030, *t* = -2.09, conflicting, β = -0.065, *SE* = 0.027, *t* = -2.35, and congruent gaze cue conditions, β = -0.109, *SE* = 0.029, *t* = -3.68, all resulted in significantly earlier first fixations on the target. These data are presented in Figure 57 below.

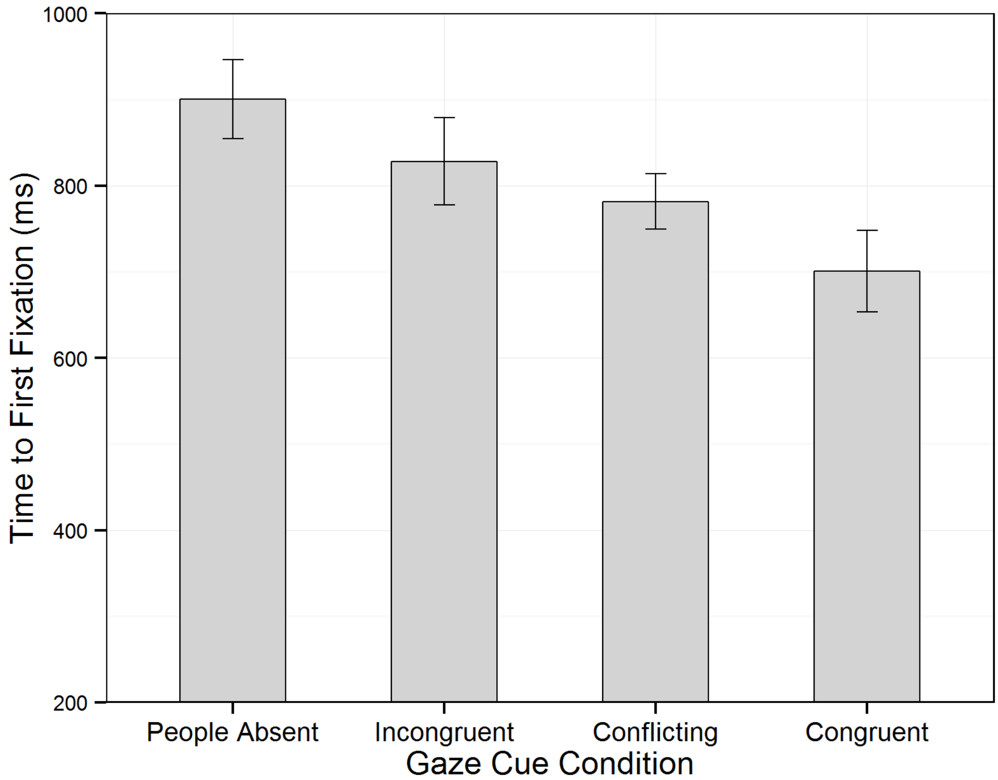

*Figure 57***.** The time to first fixation on the target (ms) from scene presentation across four gaze cue conditions. Error bars show standard error across all data samples.

Further analysis compared performance across the three people present gaze cue conditions. This showed the incongruent and conflicting gaze cue conditions resulted in almost identical times taken to first fixate the target ($t < 1$). The congruent condition produced earlier first fixations than the incongruent condition, $\beta = -0.050$, $SE = 0.026$, $t = -1.94$, which is approaching significance, and significantly earlier first fixations than the conflicting gaze cue condition, $\beta = 0.045$, $SE = 0.022$, $t = 2.03$.

Response time provides a broader measure of how long it took participants to find the target, measuring the length of time between scene presentation and the press of the trigger button on the gamepad. As in time to first fixation, person presence resulted in significantly better performance than the people absent gaze cue condition. The incongruent, $\beta = -0.107$, $SE = 0.022$, $t = -4.87$, conflicting, β = -

0.119, *SE* = 0.020, *t* = -5.92, and congruent gaze cue conditions, β = -0.107, *SE* = 0.023, *t* = -4.56, all resulted in faster response times, suggesting that participants found the target faster once people were present in the scene, as is shown in Figure 58.

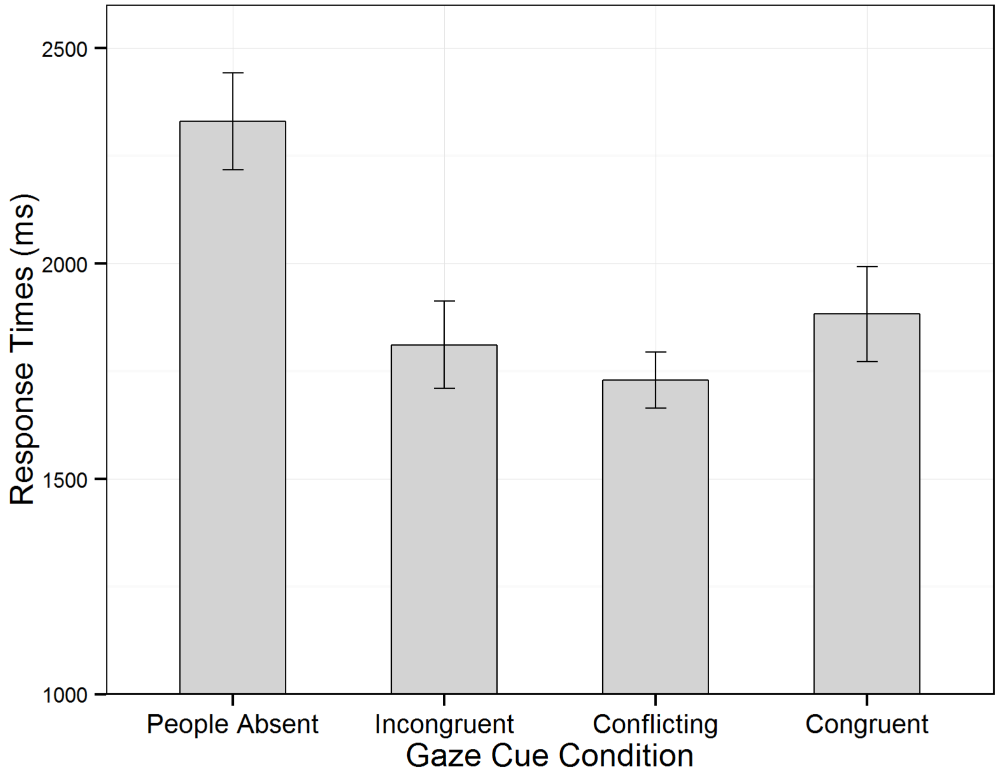

Figure 58. Response times (ms) to button press indicating successful search for the target across four gaze cue conditions. Error bars show standard error across all data samples.

As might be expected from the data presented in Figure 58, further analysis comparing the three people present gaze cue conditions showed no significant differences between them (*t*s < 1).

Scan path ratio provides a measure of search efficiency. The participants' actual scan path length is divided by the optimal scan path length; scan path ratios closer to 1 are more efficient. The scan path ratios across all four gaze cue conditions are presented below in Figure 59.

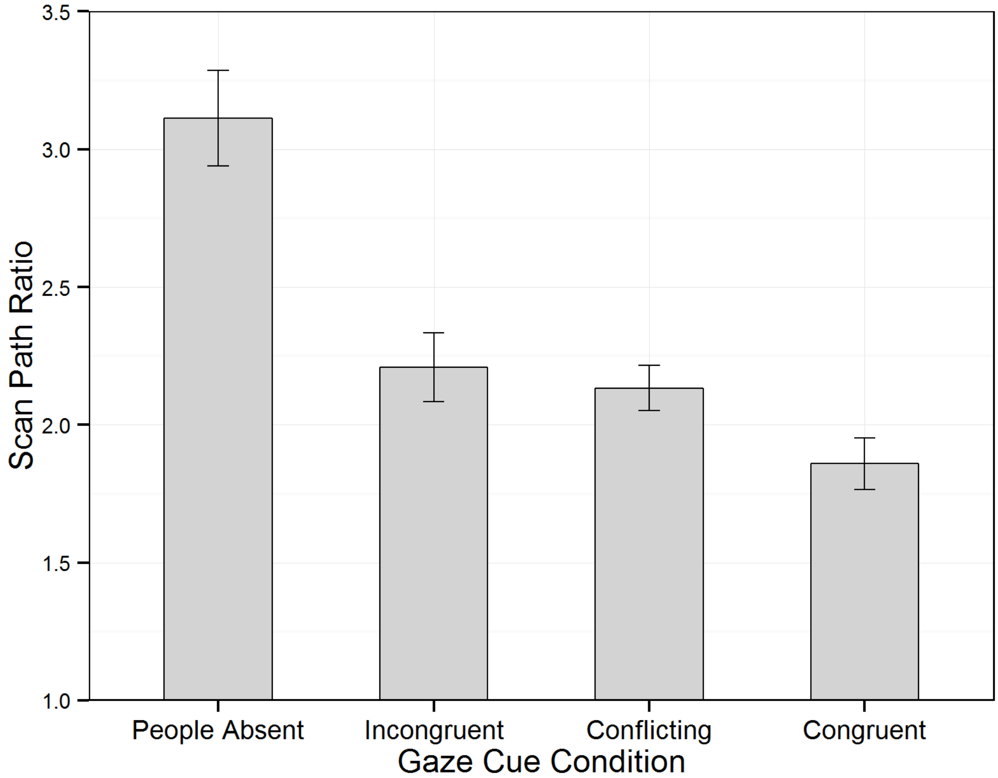

Figure 59. The scan path ratio across four gaze cue conditions. Error bars show the standard error across all data samples

As in the previous measures in the scene scanning phase, people presence dramatically proves performance within the scan path ratio measure. When compared to the people absent gaze cue condition, the incongruent, β = -0.891, SE = 0.208, t = -4.281, conflicting, β = -0.961, SE = 0.213, t = -4.514, and congruent gaze cue conditions, β = -1.241, SE = 0.234, t = -5.290, all resulted in significantly more efficient searches. Comparisons between people present gaze cue conditions showed that the congruent gaze cue condition was more efficient than both the incongruent, β = 0.384, SE = 0.181, t = 1.922, and conflicting gaze cue conditions, β = 0.281, SE = 0.153, t = 1.833, though this improvement was only approaching significance.

As a measure of actual success in locating the target, error rate present the proportion of false-positive responses where participants press the button indicating they have found the target without ever having fixated it. Significantly more errors were produced in the incongruent, $\beta = 0.128$, $SE = 0.033$, $t = 3.830$, and conflicting gaze cue conditions, $\beta = 0.111$, $SE = 0.028$, $t = 3.855$, than the people absent condition, and the congruent condition also produced a greater number of errors, $\beta = 0.059$, $SE = 0.032$, $t = 1.801$, though this was only approaching significance. Further analysis comparing people present gaze cue conditions showed a significant decrease in errors in the congruent condition as compared to the incongruent condition, $\beta = 0.069$, $SE = 0.033$, $t = 2.080$. The congruent condition also produced less errors than the conflicting gaze cue condition, $\beta = 0.052$, $SE = 0.028$, $t = 1.834$, though this was only approaching significance. There were no differences between the incongruent and conflicting gaze cue conditions ($t < 1$).

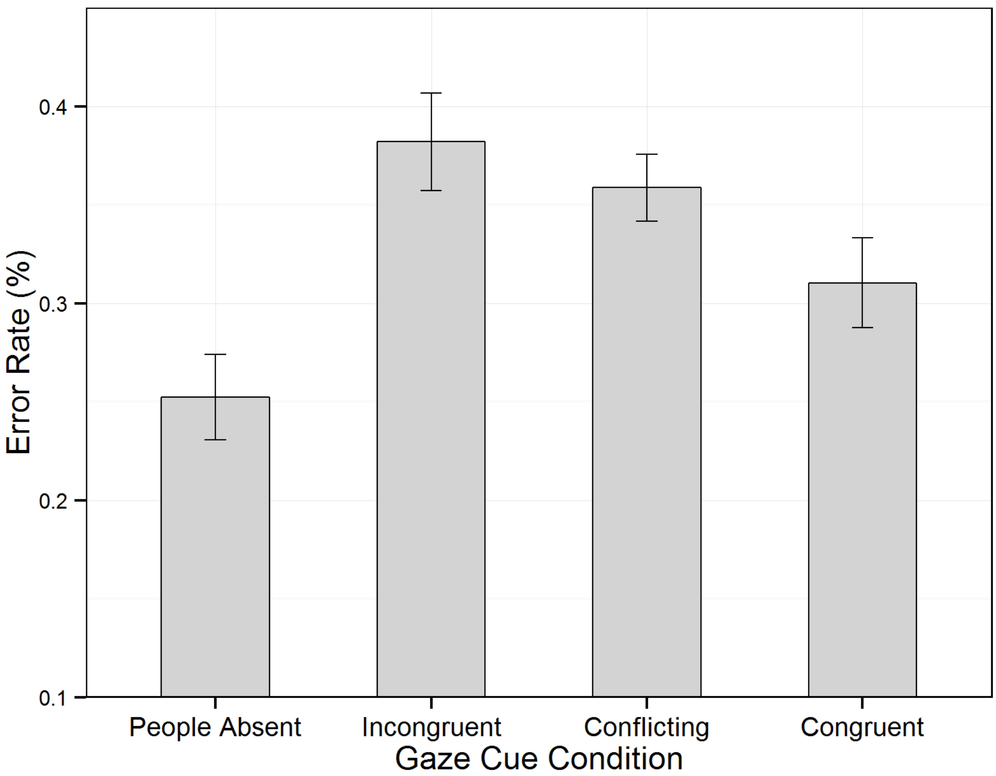

Figure 60. The proportion of false-positive responses in all trials across four gaze cue conditions. Error bars show the standard error across all data samples.

Finally, the extent to which overt fixations on the person in the scene occurred was considered using 'correct' trials; trials where no false-positive response occurred. To measure the extent of overt gaze seeking, the total number of looks towards the person in the scene was calculated, with a 'look' defined as a fixation on the head of the person. Analysis showed that across a total of 1036 trials where a person was present and a correct response occurred, 111 involved a fixation on the person's face, which accounts for 10.71% of trials.

Discussion

The studies discussed in this chapter examine the effect of two types of instruction on eye movement behaviour in a visual search task when participants are presented with multiple gaze cues simultaneously. Providing an unhelpful instruction allows comparison between previous Posner-type gaze cuing paradigms where participants are told to ignore the social cues presented. The helpful instruction addresses a previously unexplored aspect of these studies: how participants would use the social cues provided if told they may be useful to the task. Both of these studies add an additional level to our understanding of how observers' eye movement behaviours change in response to instruction in comparison to previous studies in this thesis (e.g. Chapter Three) by using multiple cues rather than a single cue.

The analyses above showed that there were some differences apparent in the way participants execute their search depending on which instruction they were given. Generally, performance in both instruction conditions followed similar patterns, but the helpful instruction elicited improved performance for more people present gaze cue conditions than an unhelpful instruction. The suggestion that people presence may be beneficial to task performance resulted in much more overt gaze-seeking. The main difference between these studies was that when given an unhelpful instruction, participants responded best in the congruent gaze cue condition, but a helpful instruction elicited more consistent performance across all people present gaze cue conditions, without any clear advantage of having a congruent gaze cue over any other type of gaze cue.

Considering the search initiation phase, both helpful and unhelpful instructions resulted in a greater proportion of first saccades directed toward the target in people present scenes than in people absent scenes, but the end point accuracy of these saccades was poorer than in the people absent condition. Only a congruent gaze cue resulted in more first saccades directed toward the target in the unhelpful instruction study, but when a helpful instruction was given even one cue directed toward the target (as in the conflicting condition) resulted in more first saccades toward the target. Across both studies, people presence resulted in poorer first saccade end point accuracy, with the incongruent condition eliciting significantly poorer accuracy than the people absent condition when an unhelpful cue was given, but both the incongruent and conflicting conditions caused deterioration in accuracy when a helpful instruction was given. These results seem to suggest that an unhelpful instruction means participants' performance is only affected when these cues are directed in the same way, whereas when a helpful instruction is given, a single cue is enough to improve or disrupt performance.

It seems performance in the search initiation phase of the unhelpful instruction study is very similar to performance described in Chapter Five, and it is therefore possible to conclude the same weighting of specific areas of the scene impact the participants' search. Rao et al. (2002) propose the first saccade is programmed by matching visual characteristics of the scene to information in the internal target template (see Malcolm & Henderson, 2009). The more detailed the target template, the greater the facilitation of search (Castelhano & Heaven, 2010; Malcolm & Henderson, 2009), and evidence from Spotorno et al. (2014) suggests that this can have an effect even in the earliest stages of search. If this is the case, it would be expected that this weighting would become even stronger when participants are told that gaze cues may be helpful for task performance. It could be tentatively suggested that the results of the helpful instruction study support such a conclusion because more first saccades are directed toward the target even when only one cue is congruent with target location. The end point accuracy data could also be viewed as supporting the hypothesis that instruction lends more weight to the contribution of gaze cues in directing search. When given an unhelpful instruction, only the incongruent condition resulted in significantly poorer accuracy than the people absent condition, whereas just one incongruent cue in the helpful instruction study was enough to cause poorer end point accuracy of the first saccade.

It may be that the perceived helpfulness of the gaze cues provided, influenced by the instructions participants are given, affects the degree to which congruency effects occur. King, Rowe and Leonards (2011) performed an experiment in which they manipulated the perceived trustworthiness of gaze cue senders and explored how this affected the degree to which participants engaged in joint attention with the gaze cue sender. Research shows faces that repeatedly cue a target object, and thus are predictive co-operative faces, are seen as more trustworthy than faces that repeatedly cue a distraction (Singer, Kiebel, Winston, Dolan & Frith, 2004). Prior to beginning the task, King et al. (2011) gave participants vignettes about the two people who would be shown giving cues throughout the task. One person was described as trustworthy, the other as untrustworthy. Participants were then asked to categorize objects as quickly as possible as either a household or garage item by pressing a key on the keyboard. At the end of the task participants gave a preference rating for the objects they had seen. King et al. (2011) found that objects that had been cued by a trustworthy sender were liked more by participants than those that had been ignored by the trustworthy sender, or which had been cued by the untrustworthy sender. Furthermore, reaction time data showed that participants were significantly slower to categorize objects that had been cued by the untrustworthy sender. These results demonstrate significant effects of the perceived trustworthiness of the gaze cue sender, and while the current study did not investigate object preference, the instruction manipulation may be somewhat similar to the manipulation of trustworthiness. If participants believe the instruction manipulation, it is possible that a helpful instruction lends more weight to the gaze cue information provided in people present scenes, which could explain why a single congruent or incongruent cue in the conflicting condition is enough to impact on participants' first eye movement in their search, whereas it is only when both cues are congruent with each other in the unhelpful instruction study that the same effects occur.

In the scene scanning phase, earlier first fixations on the target occurred when at least one cue was directed toward the target when an unhelpful instruction was given and in all people present conditions when a helpful instruction was given. All people present scenes resulted in faster response times and more efficient search than the people absent condition in both studies. In both the helpful and unhelpful instruction studies, error rates were higher than the people absent condition when at least one gaze cue was directed toward the distractor object. As in the search initiation phase, performance in the scene scanning phase of search in both instruction studies is consistent. When given a helpful instruction, even gaze cues that are incongruent to target location seem to be beneficial to participants' search, whereas in the unhelpful instruction study participants require at least one congruent gaze cue to see any significant benefit.

Recent research by Ricciardelli, Carcagno, Vallar and Bricolo (2013) suggests that this could be the result of strategic gaze following behaviour. The authors varied task instructions and gaze cue direction in a paradigm similar to that employed in Ricciardelli et al.'s (2002) study, but with the addition of distracting gaze cues to empty spatial locations. Participants were asked to make a saccade towards one of two targets that were horizontal to the centrally-presented face, which would either gaze toward the task-relevant target, gaze toward a task-irrelevant target, or toward an empty location. The authors found that while there was some evidence of reflexive orienting where participants made more errors in response to a distracting, incongruent gaze cue, participants never followed the gaze cue when it was toward an empty location. Ricciardelli et al. (2013) suggest these findings evidence that gaze following is a product of both automatic and goal-driven mechanisms. This would explain why there are still benefits derived from people presence in the unhelpful instruction study – there are still some automatic orienting mechanisms active. However, goal-driven mechanisms may only be activated in the helpful instruction condition when participants believe these cues will be useful in finding the target object. This would suggest that participants can strategically select which gaze information to attend, based on the instruction they are given, and when they are told gaze cues may be helpful greater benefit is derived from people presence than when participants believe the cues are not relevant to the task.

The measure of overt gaze-seeking shows the biggest difference between the two studies: the percentage of trials featuring a fixation on either face in the unhelpful instruction condition was only 2.08%, whereas when participants were given a helpful instruction this increased to 10.71%. Generally, performance in both instruction conditions followed similar patterns, with the helpful instruction eliciting improved performance for more people present gaze cue conditions than the unhelpful instruction, and the suggestion that people presence may be beneficial to task performance resulted in much more overt gaze-seeking. Although this percentage is still small, it shows a proportion that is five times greater that of the unhelpful gaze cue condition – a remarkable difference that can only be due to the instructions given to participants, as this is the only thing that has changed between the two studies.

Despite very little overt gaze-seeking occurring in the unhelpful instruction condition, there are still clear benefits to search derived from people presence. It is possible that participants are accessing this information covertly and using it to guide their search, even though they do not believe the cues will be helpful to the task. Returning to Downing et al.'s (2004) experiment here provides possible reasoning for these effects. Downing et al. (2004) compared the effects of predictive and non-predictive cues on observers' eye movements by presenting scenes in which a person gaze a gaze cue and a tongue-pointing cue, and the authors varied which cue was predictive of target location in a Posner-like paradigm. Participants were either told that the face was not relevant to the task, or that the target was four times more likely to occur in the un-cued location. When presented with a gaze cue participants were unable to override their reflexive response to the cue, and so still made a high proportion of eye movements to the cued location. This did not occur for the tongue cue condition where participants were less likely to fixate at the cued location. Downing et al. (2004) argue that it is this that makes gaze special: our ability to override top-down biases (knowing that the cue is unhelpful) and responding to the cue anyway, which does not occur for other stimuli. Considering this in terms of the current study, it may go some way to explain why both instruction types result in similar effects of person presence, and some effects of gaze cue congruency (albeit in different stages of search) despite participants having different top-down biases depending on what instruction they were giving. If, as Downing et al. (2004) suggest, gaze is a stimulus for which we can override our top down biases, this could account for why participants still experience benefits from people presence and congruent gaze cues, regardless of what they belief about the veracity of these cues.

The studies documented in this chapter explored the effects of instruction on participant performance during a realistic search task where participants were presented with multiple gaze cues simultaneously. The results indicated that there are strong effects of people presence across almost all measures in both studies, but effects of congruency vary depending on what participants have been told about the people in the scene. When given an unhelpful instruction, participants responded best in the congruent gaze cue condition, but a helpful instruction elicited more consistent performance across all people present gaze cue conditions, without any clear advantage of having a congruent gaze cue over any other type of gaze cue. It was suggested that a helpful instruction may add further weighting to the gaze cues provided in people present scenes, which alongside the observers' internal target template, are later used to guide search (e.g. Malcolm & Henderson, 2009; Spotorno et al., 2014). Based on findings in research by King et al. (2011), it was suggested that manipulating the perceived helpfulness of gaze cue senders could evoke similar responses in participants as manipulating perceived trustworthiness. If participants believe the instruction manipulation – as is suggested by changes in overt gaze-seeking depending on the instruction given – it is possible that a helpful instruction lends more weight to the gaze cue information provided. However, if findings were solely due to the instruction manipulation the strong benefits of people presence would not be expected in the unhelpful instruction study, but Ricciardelli et al. (2013) suggest observers can follow gaze selectively. Their gaze-following is affected by task instruction, but there is still a level of automaticity, so in the unhelpful instruction study participants are still reflexively following gaze to some degree, but this occurs more when it is strategically beneficial to do so. While performance seems to follow similar patterns in both instruction studies, this chapter does not provide a quantitative analysis of the effects of instruction. To truly ascertain whether instruction affects participant performance in the multiple-cue visual search task all three instruction conditions must compared, and this will be the focus of the next chapter.

Chapter Seven

Comparing the effects of different task instructions on observer eye movements in a visual search task featuring multiple gaze cues

Chapters Five and Six developed the novel paradigm first introduced in Chapter Two by presenting observers with two gaze cues simultaneously in the visual search task rather than one. These cues may be directed together toward the target or distractor object, or conflict with each other, with the target and distractor objects each being cued by one of the people in the scene. The effects of multiple cues on observer performance in the search task were first examined without giving any instruction other than to find the target object as quickly as possible, and the effects of different gaze cues conditions in the task are documented in Chapter Five. Instructions that made people presence salient were then introduced in Chapter Six. These instructions were designed to manipulate the perceived helpfulness of the gaze cues. Participants were either told that the gaze cues may be helpful and may help them find the target faster (a ‘helpful’ instruction), or that the people were not relevant to the current experiment and were to be ignored (an ‘unhelpful’ instruction). Chapter Six documented the effects of different gaze cues on participant performance in each of these conditions.

Unlike the two previous chapters, which examined the effects of gaze cues provided to participants, this chapter is designed to analyse the effect of instruction on performance in the task within each gaze cue condition. This allows direct comparison of the findings in each of the three studies and permits quantitative comparisons of the effects of instruction in each performance measure.

Data Analysis

The main purpose of this chapter is to examine how instructions varied the manner in which gaze cues were responded to across the three multiple gaze cue studies documented in Chapters Five and Six. Analyses in this chapter provide a quantitative analysis of the effects of instruction, and also allow for investigation into whether there are any interactions between these independent variables. To explore the influence of interactions, LMM models were run using the R statistical analysis environment (R Development Core Team, 2011). Gaze cue condition was a four-level factor, and instruction was a three-level factor, allowing three contrasts to be set up for gaze cue condition and two contrasts to be set up for instruction condition using the contrasts() function. For the gaze cue factor, the first contrast was set up to compare the person absent scenes to the person present scenes (combining across the three people present gaze cue conditions). The second contrast was set up to compare the conflicting gaze cue conditions to the gaze cue conditions in which both people cued the same object (combining the incongruent and congruent gaze cue conditions), ignoring the people absent condition. The third contrast was set up to compare the congruent and incongruent gaze cue conditions to each other, ignoring the people absent and conflicting gaze cue conditions.

For the instruction type factor, the first contrast was set up to consider whether the no instruction condition differed from the instruction conditions which make person presence salient (helpful and unhelpful instruction conditions combined). The second contrast was set up to examine whether the helpful and unhelpful instruction conditions differ from each other, ignoring the no instruction condition. The LMM was set up to consider the three contrasts for gaze cue condition and the two contrasts for instruction condition, and their interactions (thus six interaction terms describing whether each contrast in one factor depends upon each contrast in the other factor).

The results of the contrasts and LMM interactive model are presented in a similar way to the analyses of the previous chapters with *t*-values for linear models, but it should be noted that in this chapter the statistics describe whether each coded contrast is significant rather than comparisons between individual levels of a factor as in other chapters. Again, as in previous chapters, we consider any effects for which the *t*-value is greater than two as reflecting a significant effect (as in Kleigl et al., 2012).

Results

This chapter follows a similar format to that seen in Chapter Four. It allows the comparison within each gaze cue condition across all instruction types. This includes the initial study discussed in Chapter Five in which participants were given no instruction regarding the presence of people within a scene. It also includes the studies from Chapter Six, where participants were either given an instruction that suggested gaze cues may be helpful (the ‘helpful instruction’ condition), or that the gaze cues were to be ignored as they were unhelpful (the ‘unhelpful instruction’ condition).

As in previous chapters, there are some measures which require logarithmic transformation to satisfy model assumptions of normal distribution. These will be highlighted when introducing any models that required such transformations. If there were any issues with convergence, the first step was always to remove correlation parameters. In all measures where convergence was not initially achieved, the removal of correlation parameters was sufficient for convergence to occur.

First Saccade Latency

Figure 61 shows the data prior to logarithmic transformation. Analysis required this transformation to generate a normal distribution in order to satisfy model assumptions.

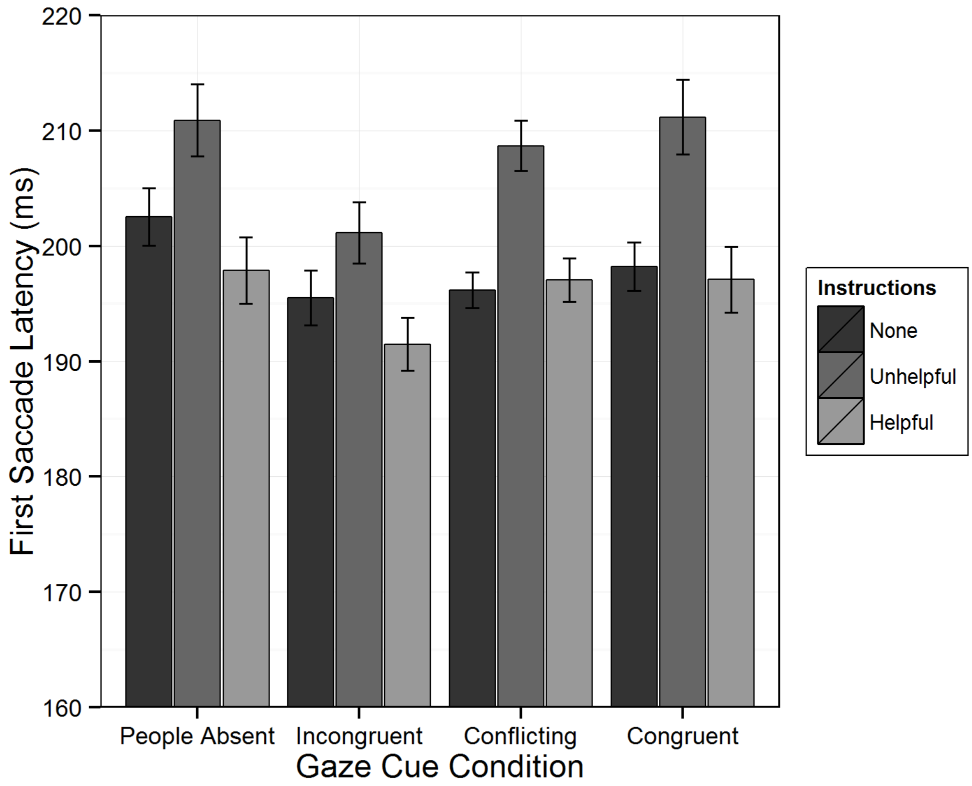

Figure 61. First saccade latency (ms) across all four gaze cue conditions, with error bars displaying standard error across all data samples. The dark grey bar represents the no instruction condition, medium grey bar represents the false instruction condition and the light grey bar represents the true instruction condition.

Contrasts of gaze cue conditions showed that people absent scenes resulted in longer first saccade latencies than people present scenes, $\beta = 0.008$, $SE = 0.003$, $t = 2.1$. The conflicting gaze cue condition resulted in first saccade latencies that were marginally longer than in gaze cue conditions where both people gazed toward the same object (congruent/incongruent), $\beta = 0.012$, $SE = 0.006$, $t = 1.9$, but this was not significant. First saccade latencies in the congruent gaze cue condition were significantly longer than in the incongruent gaze cue condition, $\beta = -0.019$, $SE = 0.008$, $t = -2.5$.

Considering the instruction conditions, there was no overall difference of first saccade latencies between instructions which made people salient (helpful/unhelpful) compared to instructions which did not make people salient (no instruction; $t < 1$). However, a helpful instruction resulted in significantly shorter first saccade latencies than those produced in the unhelpful instruction condition, $\beta = 0.024$, $SE = 0.003$, $t = 7.0$.

Analyses then examined any interactions between gaze cue conditions and instruction conditions. It was found that there was no difference in first saccade latencies produced in people present versus people absent scenes depending on whether instructions made people presence salient or not ($t < 1$), nor were there any differences between any of the three people present gaze cue conditions in these instruction conditions (ts < 1). There was no difference in first saccade latencies between people present versus people absent scenes across helpful or unhelpful instruction conditions, and no differences were observed in first saccade latencies across people present gaze cue conditions depending on whether a helpful or unhelpful instruction was given (ts < 1). This suggests no interaction between gaze cue condition and instruction condition in this measure.

First Saccade Direction

Figure 62 below shows the proportion of first saccades directed toward the target across all three instruction types and all four gaze cue conditions.

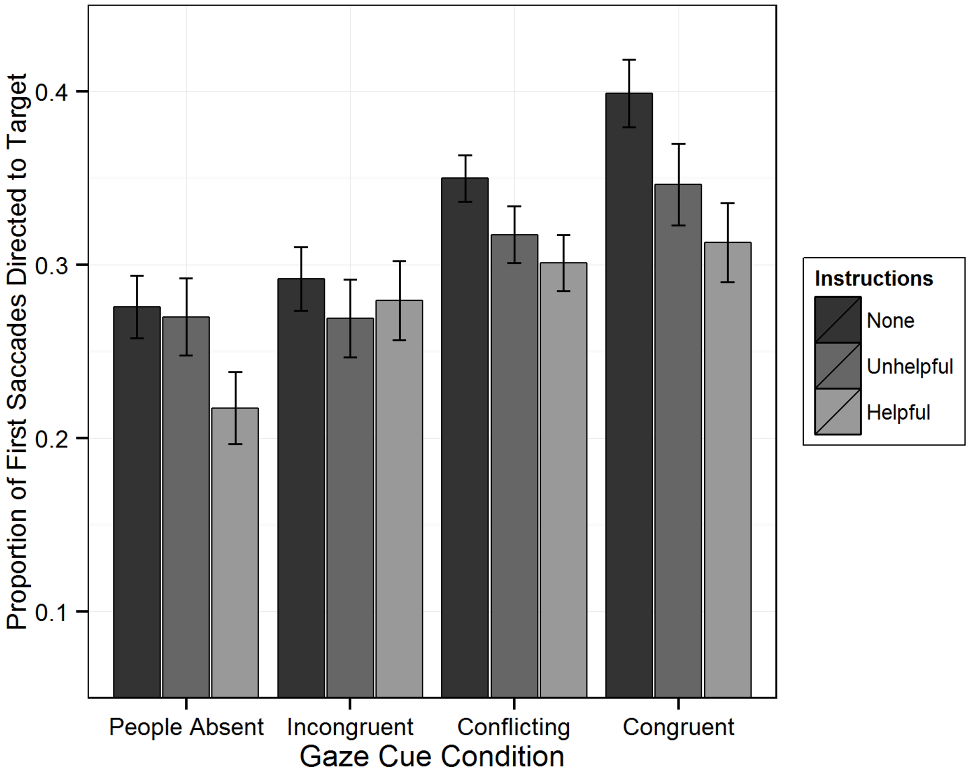

Figure 62. The proportion of first saccades directed toward the target across all four gaze cue conditions, with error bars displaying standard error across all data samples.

Contrasts of gaze cue conditions showed people absent scenes resulted in a significantly lower proportion of first saccades directed toward the target than people present scenes, β = -0.063, *SE* = 0.015, *t* = -4.160. The conflicting gaze cue condition resulted in a proportion of first saccades directed toward the target that was significantly different from conditions where both people cued the same object, β = 0.106, *SE* = 0.031, *t* = 3.347. Follow up contrasts showed that the proportion of first saccades directed toward the target in the conflicting gaze cue condition was greater than the incongruent gaze cue condition, β = -0.066, *SE* = -0.022, *t* = -2.938, and lower than the congruent gaze cue condition, β = 0.070, *SE* = 0.020, *t* = 3.347. Contrasts also showed that the incongruent gaze cue condition resulted in a much smaller proportion of first saccades being launched in the direction of the target compared to the congruent gaze cue condition, β = -0.136, *SE* = 0.037, *t* = -3.598.

Contrasts of instruction conditions demonstrated that there was a significantly higher proportion of first saccades directed toward the target when the instruction made no reference to people presence, compared to instructions where people presence was made salient, β = 0.031, *SE* = 0.013, *t* = 2.336, However, there was no difference between the proportion of first saccades directed toward the target in the helpful versus unhelpful instruction conditions, β = 0.022, *SE* = 0.015, *t* = 1.527.

The interactions between gaze cue condition and instruction condition were examined in further analyses. It was found that there was no difference in the proportion of first saccades directed toward the target in people present versus people absent scenes depending on whether instructions made people presence salient or not ($t < 1$). Similarly, there were no differences in proportion of first saccades directed toward the target between the conflicting and congruent or incongruent gaze cue conditions, β = 0.074, *SE* = 0.059, *t* = 1.258, or the congruent and incongruent conditions, β = -0.111, *SE* = 0.070, *t* = -1.571, depending on whether instructions made people presence salient or not. There was no difference in the proportion of first saccades directed toward the target across people absent versus people present scenes in helpful or unhelpful instruction conditions, β = 0.039, *SE* = 0.036, *t* = 1.087, and there were no differences in proportion of first saccades directed toward the target in any of the three people present gaze cue conditions across helpful or unhelpful instruction condition (*t*s < 1). This suggests no interaction between gaze cue condition and instruction condition in this measure.

End Point Accuracy of the First Saccade

The end point accuracy of the first saccade – that is, how close the saccade brought the eyes to the centre of the target – is the final measure of the search initiation phase. The data across all four gaze cue conditions are presented below in Figure 63.

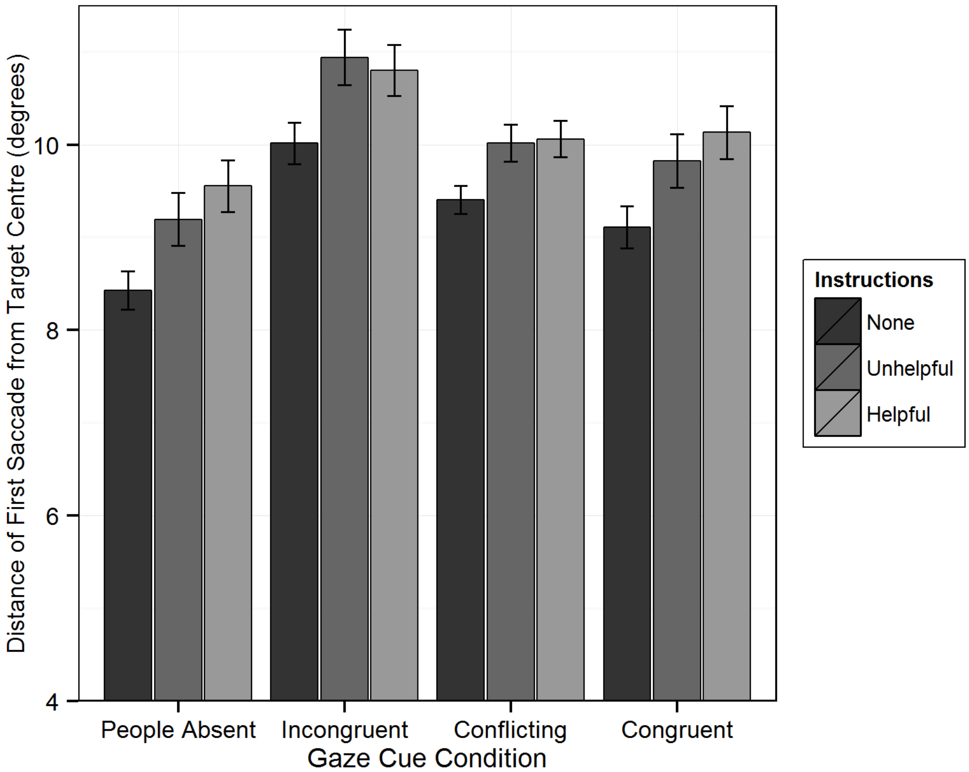

Figure 63. The distance of the landing point of the first saccade from the centre of the target boundary box (in degrees of visual angle) as a measure of end point accuracy across four gaze cue conditions. Error bars show standard error across all data samples.

Contrasts of gaze cue conditions showed end point accuracy was significantly better in people absent scenes compared to people present scenes, β = -1.045, *SE* = 0.177, *t* = -5.893. The conflicting gaze cue condition resulted in significantly different end point accuracy than gaze cue conditions where both people cued the same object, β = -1.259, *SE* = 0.433, *t* = -2.904. Follow up contrasts showed end point accuracy in the conflicting condition was better than the incongruent gaze cue condition, $\beta = -0.920$, $SE = 0.228$, $t = -4.030$, but poorer than end point accuracy in the congruent gaze cue condition, $\beta = 0.723$, $SE = 0.225$, $t = 3.208$. Contrasts showed end point accuracy was significantly poorer in the incongruent gaze cue condition compared to the congruent gaze cue condition, $\beta = 1.783$, $SE = 0.476$, $t = 3.741$.

Considering the instruction conditions, contrasts demonstrated that end point accuracy was significantly poorer in instruction conditions which made people presence salient compared to the no instruction condition, $\beta = -0.644$, $SE = 0.134$, $t = -4.785$, but there was no difference in end point accuracy between the helpful and unhelpful instruction conditions ($t < 1$).

Further analyses of the interactions between gaze cue conditions and instructions conditions demonstrated that end point accuracy in people present versus people absent scenes was unaffected by whether instructions made people presence salient or not ($t < 1$), nor did the salience of people presence in instructions given to participants impact end point accuracy across any of the three people present gaze cue conditions (*t*s < 1). Having a helpful or unhelpful instruction did not affect end point accuracy in people present versus people absent scenes ($t < 1$), nor did it affect end point accuracy across any of the three people present gaze cue conditions (*t*s < 1). This suggests no interaction between gaze cue condition and instruction condition in this measure.

Time to First Fixation on the Target

Figure 64 displays the time (in ms) taken for the participant to first fixate on the target object across all four gaze cue conditions and three instruction conditions. Note that Figure 64 presents the data untransformed; data underwent logarithmic transformation for analyses.

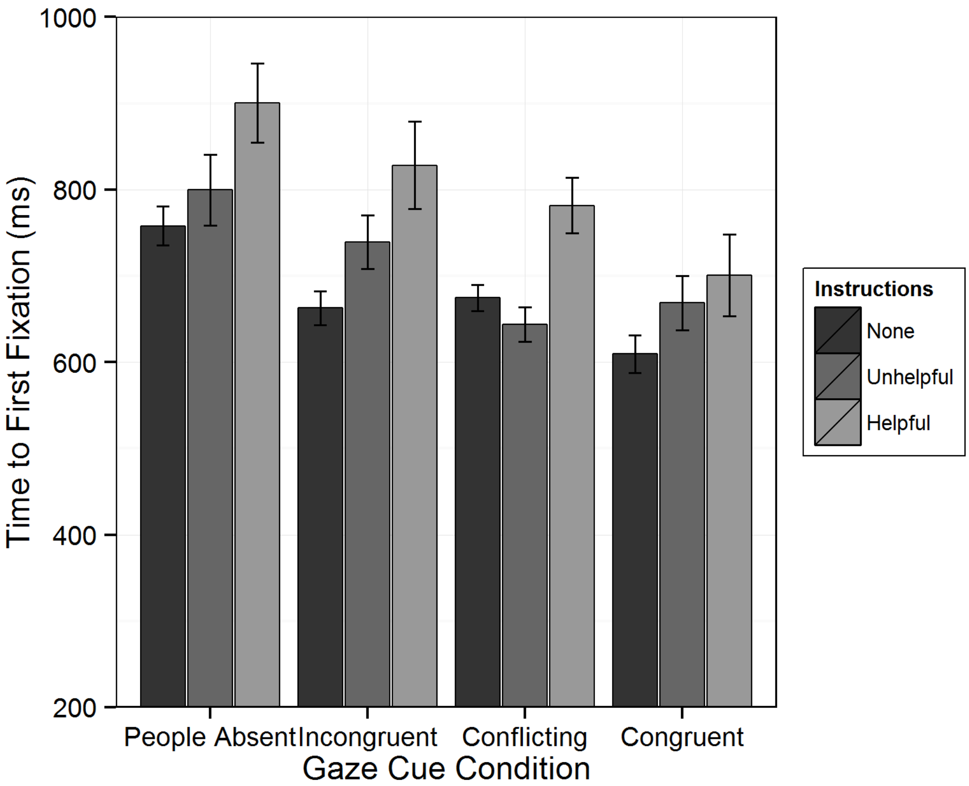

Figure 64. The time to first fixation on the target (ms) from scene presentation across four gaze cue conditions, where error bars show standard error across all data samples.

Contrasts of gaze cue conditions showed the time to first fixation was significant faster in people present scenes compared to people absent scenes, β = 0.061, *SE* = 0.010, *t* = 5.73. The conflicting gaze cue condition resulted in significantly different first fixations to gaze cue conditions where both people cued the same object, β = -0.081, *SE* = 0.026, *t* = -3.07. Time to first fixation was shorter in the conflicting condition than the incongruent condition, β = -0.046, *SE* = 0.014, *t*

= -3.18. However, first fixations on the target occurred later in the conflicting gaze cue condition than in the congruent gaze cue condition, β = 0.054, *SE* = 0.017, *t* = 3.07. First fixations on the target were significantly faster in the congruent gaze cue condition compared to the incongruent gaze cue condition, β = 0.100, *SE* = 0.028, *t* = 3.55.

Contrasts of the instruction conditions demonstrated that there were faster first fixations on the target when instructions did not make reference to people presence compared to instruction conditions which made people presence salient, β = -0.028, *SE* = 0.009, *t* = -2.90. An unhelpful instruction resulted in faster first fixations than a helpful instruction, β = -0.032, *SE* = 0.011, *t* = -2.86.

Further analyses of the interactions between gaze cue conditions and instructions conditions demonstrated that the time to first fixation on the target in people present versus people absent scenes was unaffected by whether instructions made people presence salient or not ($t < 1$), nor did the salience of people presence in instructions given to participants impact end point accuracy across any of the three people present gaze cue conditions (*t*s < 1). Having a helpful or unhelpful instruction did not affect end point accuracy in people present versus people absent scenes, β = -0.033, *SE* = 0.027, *t* = -1.22, nor did it affect end point accuracy across any of the three people present gaze cue conditions (*t*s < 1). This suggests no interaction between gaze cue condition and instruction condition in this measure.

Response Time

Figure 65 below presents the response time data across all four gaze cue conditions and three instruction conditions untransformed. Data underwent logarithmic transformation for analyses to be conducted.

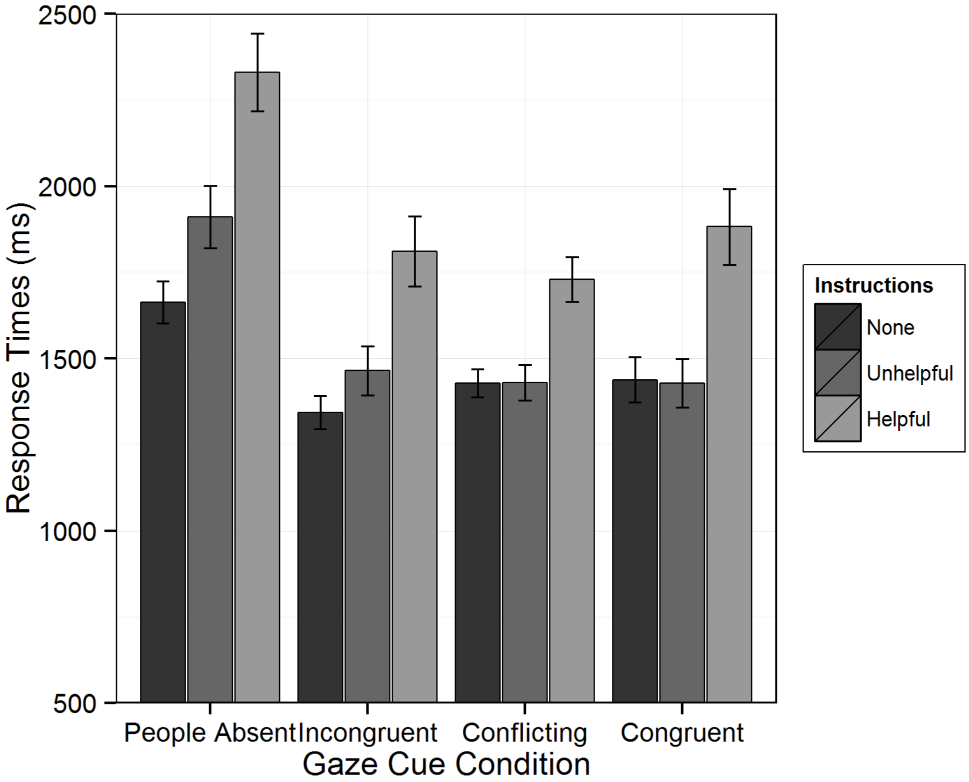

Figure 65. Response times (ms) to button press indicating successful search for the target across four gaze cue conditions, with error bars showing standard error across all data samples.

While contrasts of gaze cue condition demonstrated significantly slower response times in people absent scenes compared to people present scenes, β = 0.100, *SE* = 0.008, t = 11.51, there was no difference in response times between the conflicting condition and gaze cue conditions where both people cued the same object or between the congruent and incongruent gaze cue conditions (ts < 1).

Contrasts of instruction condition showed that having an instruction that made people presence salient resulted in significantly slower response times than instructions which did not make reference to people presence, β = -0.072, *SE* = 0.009, *t* = -7.87. Further contrasts showed a helpful instruction resulted in significantly slower response times than in the unhelpful instruction condition, β = -0.070, *SE* = 0.009, *t* = -7.21.

Investigation of the interaction between gaze cue condition and instruction condition showed that response times were affected in people present versus people absent scenes depending on whether the instruction had made people presence salient or not, β = -0.040, *SE* = 0.016, *t* = -2.48. However the salience of people presence in instruction did not result in any difference in response times in the conflicting condition compared to gaze cue conditions in which both people cued the same object, nor between the congruent and incongruent gaze cue conditions (*t*s < 0.5). There were no differences in response times in people present versus people absent scenes when a helpful or unhelpful instruction was given (*t* < 0.5), and a helpful or unhelpful instruction did not affect response times across any of the three people present gaze cue conditions (*t*s < 1). This suggests that the interaction between salience of people presence in the instruction and response times in people present versus people absent scenes is driven by the significantly higher response times which occur when people are absent from the scene.

Scan Path Ratio

The scan path ratios for each instruction condition across all four gaze cue conditions can be seen below in Figure 66.

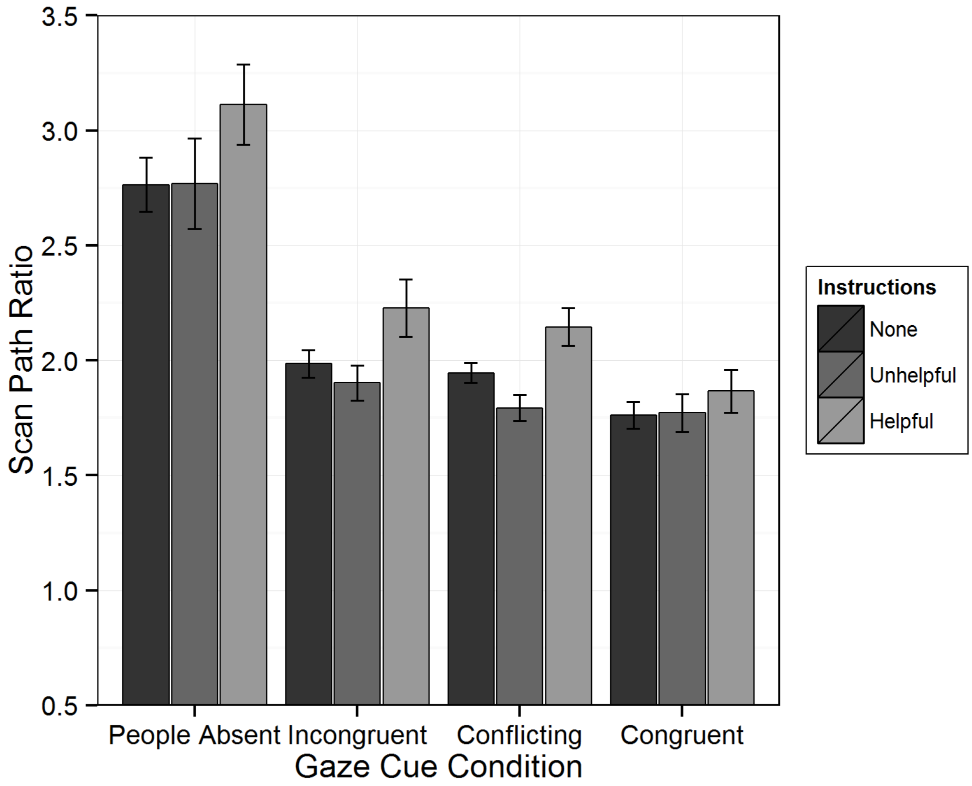

Figure 66. The scan path ratio across four gaze cue conditions with error bars showing standard error across all data samples.

Contrasts of gaze cue conditions showed that scan path ratio was significantly closer to one in people present scenes compared to people absent scenes, $\beta = 1.040$, $SE = 0.105$, $t = 9.90$. The conflicting condition resulted in scan path ratios that were significantly different from gaze cue conditions in which both people cued the same object, $\beta = -0.433$, $SE = 0.145$, $t = -2.99$. Follow up contrasts showed the conflicting gaze cue condition elicited searches that were more efficient than those in the incongruent condition, $\beta = -0.206$, $SE = 0.096$, $t = -2.13$, but less efficient than those in the congruent gaze cue condition, $\beta = 0.264$, $SE = 0.094$, $t = 2.80$. Search efficiency was greater in the congruent gaze cue condition compared to the incongruent gaze cue condition, $\beta = 0.490$, $SE = 0.172$, $t = 2.85$.

Considering instruction conditions, contrasts showed that scan path ratios were closer to one when instructions did not make reference to people presence in the scene compared to conditions which made people presence salient, $\beta = -0.146$, $SE = 0.060$, $t = -2.40$. A helpful instruction resulted in a much less efficient search than the unhelpful instruction condition, $\beta = -0.275$, $SE = 0.076$, $t = -2.40$.

Examination of the interactions between gaze cue condition and instruction condition showed that there was some difference in scan path ration in people present versus people absent scenes when instructions made people presence salient compared to when they made no reference to people presence, $\beta = -0.278$, $SE = 0.146$, $t = -1.90$, but this was not significant. There was no difference in search efficiency between the congruent condition and gaze cue conditions where both people cued the same object depending on whether instructions made people presence salient ($t < 0.5$). Similarly, the salience of people presence in the instructions did not affect scan path ratio in the incongruent gaze cue condition compared to the congruent gaze cue condition ($t < 0.5$). Having a helpful or unhelpful instruction did not impact search efficiency in people present versus people absent scenes ($t < 0.5$). Scan path ratios in the conflicting gaze cue condition compared to conditions where both people cued the same object were unaffected by having a helpful or unhelpful instruction, $\beta = 0.496$, $SE = 0.380$, $t = 1.31$, and there was no difference in scan path ratio between the incongruent and congruent gaze cue conditions depending on whether a helpful or unhelpful instruction was given, $\beta = -0.461$, $SE = 0.460$, $t = -1.00$. This suggests no interaction between gaze cue condition and instruction condition in this measure.

Error Rate

Error rate data – that is, measures of false-positive responses made by participants during search – is presented in Figure 67 below. As might be expected from the data presented in Figure 67, there were considerable differences in error rate responses across the instruction conditions in all gaze cue conditions.

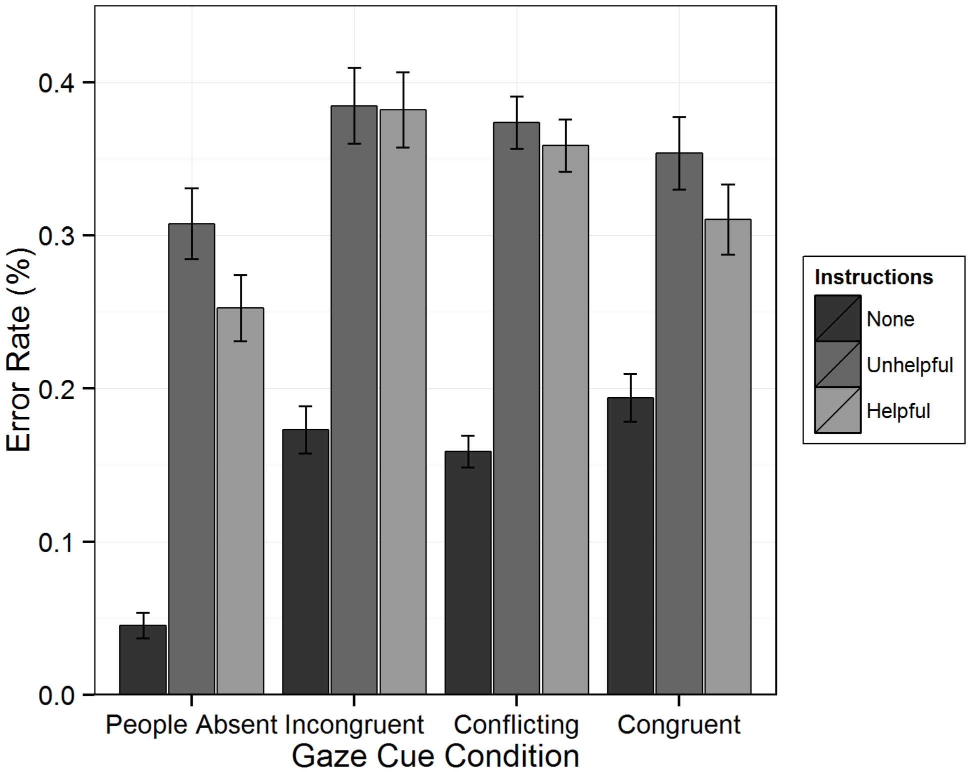

Figure 67. The error rate across four gaze cue conditions, shown as a percentage of trials in which participants made a false-positive response. Error bars show the standard error across all data samples.

Contrasts showed that there were significantly more errors in people present scenes compared to people absent scenes, β = -0.097, *SE* = 0.013, *t* = -7.524. Though there were some differences in performance between the conflicting condition and gaze cue conditions where both people cued the same target, β = -0.033, *SE* = 0.031, *t* = -1.038, and between the congruent and incongruent gaze cue conditions, β = 0.047, *SE* = 0.035, *t* = 1.323, these differences were not significant.

Error rate was significantly higher when instructions given to participants made reference to people presence compared to the no instruction condition, β = -0.195, *SE* = 0.010, *t* = -17.957. There were more errors in the unhelpful instruction condition than the helpful instruction condition, β = 0.028, *SE* = 0.013, *t* = 2.069.

Examination of the interaction between gaze cue condition and instruction condition showed that while there was some difference in error rate in people present versus people absent scenes depending on whether instructions made reference to people presence or not, β = -0.048, *SE* = 0.025, *t* = -1.891, this was not significant. However, there was a significant difference in error rate in the conflicting condition compared to gaze cue conditions where both people cued the same object depending on whether instructions made people presence salient or not, β = 0.138, *SE* = 0.056, *t* = 2.453. Error rates in the incongruent versus congruent gaze cue conditions were significantly affected by instructions making reference to people presence, β = -0.138, *SE* = 0.066, *t* = -2.079. However, when instructions that make people salient are broken down into helpful and unhelpful instruction conditions, these interactions disappear. Having a helpful or unhelpful instruction does not affect error rate in people present versus people absent scenes, β = 0.034, *SE* = 0.033, *t* = 1.027, in the conflicting condition versus gaze cue conditions where both people cue the same object, nor in the congruent versus incongruent gaze cue conditions (*t*s < 1). This suggests that when comparing the people absent scenes to all people present scenes, making people presence salient in the instruction results in significantly higher error rates than when no mention is made of people presence in the instruction. However, there is no effect of suggesting whether these people may be helpful or unhelpful.

Overt Gaze-Seeking

As a final measure of scene scanning the proportion of fixations on either person in the scene was examined across the three people present gaze cue conditions. A fixation on the face of either person was defined as a fixation that fell within the boundary box around either person's head; the fixation did not have to land on the eyes. Any fixations within this facial region would be examples of overt gaze following behaviour. The number of fixations on the face of either person in the scene are displayed below in Figure 68.

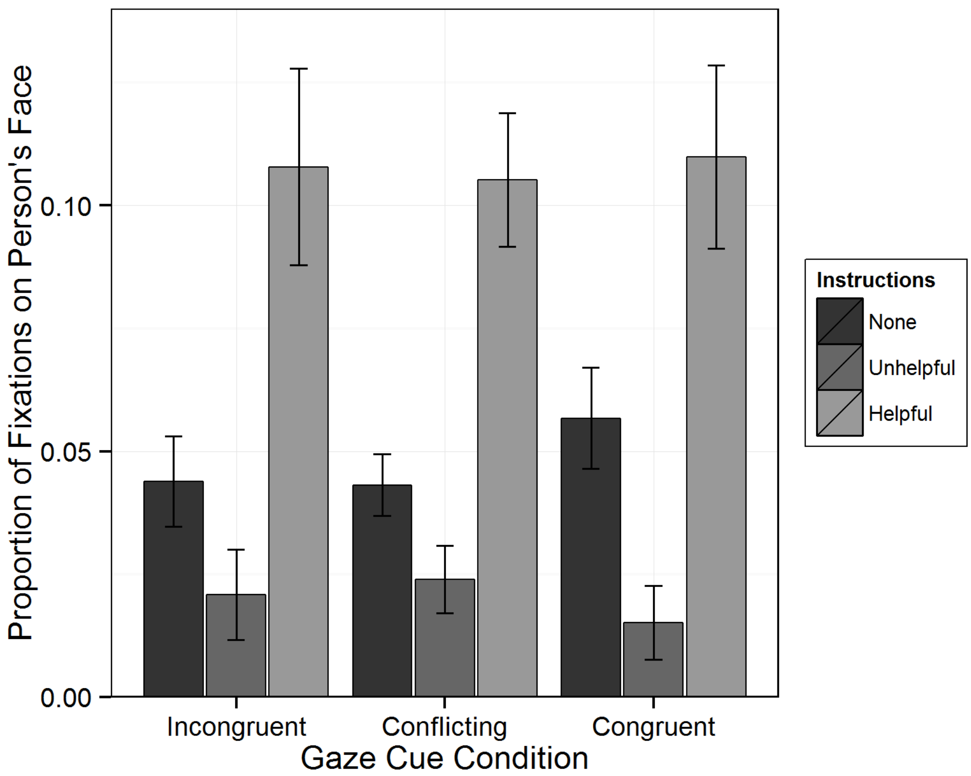

Figure 68. The proportion of fixations on the face of either individual present in the scene across the people present gaze cue conditions where error bars show standard error across all data samples.

Contrasts showed that there was no difference in the rate of overt gaze-seeking between the incongruent gaze cue condition and the conflicting or congruent gaze cue conditions (*t*s < 1). There was no difference in overt gaze-seeking in instructions which made reference to people presence compared to the no instruction condition ($t < 1$), but a helpful instruction did result in significantly more overt gaze-seeking than an unhelpful instruction, $\beta = -0.086$, $SE = 0.019$, $t = -4.381$.

Further analyses of the interactions between gaze cue conditions and instruction conditions showed that instructions that made people presence salient did not result in different levels of overt gaze-seeking in the conflicting gaze cue condition compared to conditions in which both people cued the same target ($t < 1$). Similarly, instructions which did or did not make people presence salient did not affect levels of overt gaze-seeking in congruent versus incongruent gaze cue trials ($t < 1$). There was no difference in overt gaze-seeking across any of the three people present gaze cue conditions when a helpful instruction was given compared to an unhelpful instruction (*t*s < 1). This suggests no interaction between gaze cue condition and instruction condition in this measure.

Discussion

This chapter directly compared the previous three experiments discussed in Chapters Five and Six that presented participants with multiple gaze cues in the visual search task, and each featured different task instructions regarding the purpose of people presence in the scenes. Participants performed the visual search task in one of three instruction conditions: participants were told nothing about people presence in the scene (no instruction; Chapter Five), that the people were to be ignored (unhelpful instruction), or that the people may help in finding the target object faster (helpful instruction; both Chapter Six). Unlike the previous two chapters which examined the effects of gaze cue conditions on participants' performance, this chapter compares performance across instruction conditions. This allows quantitative comparison of the effects of instruction so that it can be determined whether each instruction condition affects observer eye movement behaviour and if so, in what way. This chapter also explores whether there is any interaction between gaze cue condition and instruction condition across all eight performance measures.

Generally, these measures indicate that there are clear effects of instruction on participants' behaviour in the task. In most measures, any instruction regarding people presence causes results consistent with deterioration in search performance, and this tends to be strongest when participants are given an instruction that suggests the gaze cues may be helpful in finding the target object. It is only in the first saccade latency measure, which examines pre-saccadic launch processing, that a helpful instruction resulted in faster responses than an unhelpful instruction. In addition, one result that particularly stands out is the measure of overt gaze seeking, which showed that when people believe a cue may be helpful they fixate considerably more on the peoples' faces than when there is no instruction regarding people presence or when they believe gaze cues are not helpful. This final measure certainly suggests that the manipulation was successful and that participants believed the instruction that was given to them.

Analyses show that first saccade latency is somewhat atypical amongst search initiation measures as is it the only one in which an instruction that suggests gaze cues may be helpful facilitates performance compared to the unhelpful instruction condition. Contrasts of gaze cue condition showed all people present gaze cue conditions had shorter first saccade latencies than the people absent scenes, with the shortest latencies in the congruent gaze cue condition. Of the people present gaze cue conditions, the incongruent condition produced the longest latencies, with the conflicting gaze cue condition falling in the middle, though it was not significantly different from either of the other people present conditions. There were no apparent differences in first saccade latency when no instruction was contrasted with instructions which made people presence salient. It is likely that because latencies were similar in the no instruction and unhelpful instruction conditions the first contrast found no difference when the unhelpful and helpful instructions were grouped together. Further contrasts showed the helpful instruction condition resulted in significantly shorter first saccade latencies than the unhelpful instruction condition. However, as stated, these results are unusual amongst search initiation measures. In measures of first saccade direction and the end point accuracy of the first saccade, participant performance when given an instruction that made people presence salient was worse than when no instruction was given about people presence, but there were no further significant differences between the helpful and unhelpful instruction conditions. Across all three measures of search initiation there were no interactions between gaze cue condition and instruction condition, suggesting that while there are effects of gaze cue condition and instruction condition on participants' performance, the effects of gaze cue condition and instruction condition occur independently of the other variable.

While the effects of instruction in this phase of search are unexpected, they do evidence that that instruction manipulation has impacted on participants' performance. This corroborates previous findings that demonstrate instructions affect observers' eye movement behaviour. That task demands, directed by instructions given to participants, can affect participant behaviour is a long-established finding within visual search literature. For example, Einhäuser, Rutihauser and Koch (2008) demonstrated that when viewing photographs of outdoor scenes, instructions to freely-view the scenes resulted in fixation bias to the high-contrast area of the scene, but instructions to search for a bulls-eye target that was equally likely to appear on either side of the scene resulted in focused search for the target. The authors present this as evidence that task demands can override sensory-driven information. It has been suggested that individuals can regulate their attention to be sensitive to different stimuli depending on the task at hand, improving their efficiency in finding their target (Folk, Remington & Johnstone, 1992). Recent research has even shown that when eye movements are analysed in a computational model, the task that was given to participants can be predicted from the eye movements they make, in what Haji-Abohhassani and Clark (2014) dub 'an inverse Yarbus process'.

Considering the methodology used in the current studies, it is also useful to return to the investigation of task demands on gaze orienting conducted by Itier et al. (2007). They conducted a Posner-like task in which used the same stimuli across two tasks. The stimuli featured a face presented in front or three-quarter view with direct or averted gaze. Participants were asked either to determine the direction toward which the eyes were pointing (gaze task) or which way the head was pointing (head task). Itier et al. (2007) found that in the gaze discrimination task approximately 90% of first saccades were directed toward the eye region of face, but this fell to around just 50% in the head direction discrimination task. The authors posited that while their results supported research suggesting eyes are a preferred stimulus (e.g. Henderson, Williams & Falk, 2005; Yarbus, 1967); they also demonstrated that the orienting response to gaze was not a reflexive mechanism. Rather, gaze seeking and following is an endogenous top-down process in which seeks to fulfil task demands by selecting appropriate stimuli. Considering these findings in terms of the analyses presented above, Itier et al.'s (2007) study provides strong evidence to support the effects task instruction has on participant performance, thus giving an evidence base from which the findings of the current analyses can build. Itier et al. (2007) were able to demonstrate a significant drop in the number of fixations on the eyes of the face simply by changing the task to a head orientation discrimination task rather than an eye orientation discrimination task. Otherwise, the stimuli presented to participants were the same, with the only difference being a manipulation in what participants perceived as the important element of the task. The studies described in this chapter do the same. Just like Itier et al. (2007), these studies change what is perceived to be important to the task by giving participants different information about scene content whilst presenting the same stimuli, and both result in different participant behaviour depending on the instruction condition. Automatic orienting to the gazed-at location would predict no effect of instruction, but since effects of instruction are evident in the current studies, these findings offer evidence that automatic orienting cannot wholly account for the findings of these studies.

Effects of instruction on performance continue in the scene scanning phase of search. Considering the time taken to first fixate on the target, performance was worse in instruction conditions which made people presence salient compared to the no instruction condition, and the longest time to first fixation occurred in the helpful instruction condition. For response times, a helpful instruction again resulted in the worst performance of all instruction conditions, with both instruction conditions that made people presence salient resulting in slower response times than the no instruction condition. Scan path ratio – the measure of search efficiency – followed the same pattern of results, with the no instruction condition eliciting the most efficient searches when contrasted with instruction conditions that made people presence salient, and less efficient searches in the helpful instruction condition compared to the unhelpful instruction condition. While there were no interactions between gaze cue condition and instruction condition in the time to first fixation or scan path ratio measures, response time did see a significant interaction that suggests instructions that made people presence salient compared to those that did not had a significant effect on response times in people absent versus people present scenes. Considering the contrasts of gaze cue condition, which show very large differences in response times in people absent scenes compared to people present scenes, this is unsurprising. However, there were no differences in response times between people present conditions, which would suggest instruction only has an effect at a very basic level – whether people are present in the scene or not – but has no impact depending on the types of gaze cues these people are providing.

As a measure of overall success, error rate determines the proportion of trials in which participants press the trigger button to indicate completed search when no fixation on the target has occurred. In this measure there was a significantly higher proportion of errors in people present scenes compared to people absent scenes. Any instruction which made people presence salient resulted in more errors, with the unhelpful instruction condition producing a greater proportion of errors than the helpful instruction condition. Analyses of interactions between gaze cue condition and instruction condition found significant differences in error rate in the conflicting condition compared to gaze cue conditions in which both people cued the same object, and in the incongruent versus congruent gaze cue conditions, depending on whether instructions made reference to people presence or not. However, there were no differences in these gaze cue conditions when the helpful and unhelpful instruction conditions were compared. This would suggest that it is simply the mention of people within the instruction that elicits a difference in performance across people present gaze cue conditions, rather than what the instruction conveys about the purpose of these people within the scene.

Overt gaze-seeking demonstrated consistent effects of instruction across all people present gaze cue conditions: an unhelpful instruction produced the lowest proportion of fixations on either person’s face with almost half the fixations of the ‘baseline’ no instruction condition. A helpful instruction, however, resulted in a significantly higher proportion of overt gaze-seeking – almost double the proportion of fixations in the no instruction condition. This effect provides the strongest evidence that the participants believed the instruction manipulation. However, there were no interactions between gaze cue condition and instruction condition in this measure, which suggests the effects of gaze cue condition and instruction on overt gaze-seeking occur independently of one another.

As discussed above, automatic orienting would suggest no effects of instruction, but because there are clear differences in participants’ performance across instruction conditions, automatic orienting cannot wholly account for the results found within this chapter. However, a theory of selectively attending scene content cannot fully account for these findings either. If participants are selectively attending information they believe to be helpful to their performance in the task, why does a ‘helpful’ instruction often result in the poorest performance of all three instruction conditions? The answer may come from research conducted by Fletcher-

Watson et al. (2008). These authors conducted a task in which participants were presented with two images side by side. One of these images contained a person, and the other did not. In the first half of their experiment participants were asked to freely view the scenes, and in the second half they were given a task related to their viewing: they were asked to identify the gender of the person in the scene. Their study is particularly relevant to two of the findings discussed above. While the person in the person-present half of the scene was always preferentially fixated over other stimuli, Fletcher-Watson et al. (2008) reported a significant increase in the proportion of fixations on the person when they were made salient by task instruction, which also changed how participants looked at the person. In the gender-discrimination task 40% of the total time looking at the person involved looking at their body, compared to almost no fixations on the body in the free viewing task. This is clear evidence that task instruction changes how observers look at a person (or people) in a scene. The results in the overt gaze-seeking measure detailed above clearly support these findings. When given a helpful instruction participants were significantly more likely to fixate on the faces of the people in the scene than both the no instruction and unhelpful instruction conditions; therefore the current studies also demonstrate that different instructions regarding people presence change the degree to which overt gaze-seeking behaviours occur.

Fletcher-Watson et al.'s (2008) study also provides an explanation for why – despite the general facilitation in performance as a result of person presence – instructions that made people presence salient tended to result in poorer performance in the search task compared to the no instruction condition. To answer the secondary question of whether objects cued by the gaze of the person in the scene would be the focus of a greater number of fixations than other objects, Fletcher-

Watson et al. (2008) mapped a ‘viewing cone’ on to scenes, which essentially created a boundary for the area cued by the person’s gaze. In the free viewing condition participants fixated on the object being cued by the person in the scene significantly more than would be predicted by a random viewing pattern. However, this effect was not evident in the gender-discrimination task. The authors suggest that the results of the free viewing condition support the possibility that gaze following can occur from looking at static scenes, and that when gaze following was not relevant to the task at hand in the gender-discrimination task, fixations on the object being cued become fewer. Fletcher-Watson et al.’s (2008) instruction in the gender-discrimination task made gaze irrelevant to the task, which corresponds with the unhelpful instruction condition in the current studies. Their results would suggest that overt gaze-seeking behaviour depends on the instruction given to participants: an unhelpful instruction that makes gaze cues irrelevant to the task should hinder performance, but an instruction that suggests gaze cues are helpful facilitates performance. Considering the analyses of the studies discussed in this chapter, Fletcher-Watson et al.’s (2008) results may explain why performance across visual search measures become worse in the helpful instruction condition. The stimuli presented to participants are the same across all instruction conditions, and feature spatially uninformative cues. However, in every people present scene at least one object is cued by the people in the scene. If participants believe that these gaze cues are unhelpful their search behaviour continues as normal, which would explain why often the no instruction and unhelpful instruction conditions have similar levels of performance. Conversely, if participants are told that the gaze cues provided may be helpful in finding the target object, a greater level of importance is attributed to these cues and therefore more fixations occur within what Fletcher-

Watson et al. (2008) dubbed the ‘viewing cone’ of each person in the scene. Overall the cues are not actually helpful for completing the task – they are equally likely to cue a distractor as the target – allocating attention within the peoples’ viewing cones within the scene is essentially wasting resources. Given that the increase in fixations on either person’s face is largest in the helpful instruction condition, this would suggest that participants are indeed allocating more attention within the viewing cone cued by the people in the scene.

Within gaze cue conditions the cues are differentially informative, so it would be logical to predict that search performance should be best in the congruent gaze cue condition when participants are given a helpful instruction, but this is never the case. It is possible that task instructions suggesting the gaze cues provided are helpful – when they are not – may temporarily create a bias that undermines their ability to successfully complete the task. Folk et al. (1992) described how observers could generate a set of search parameters, which may include stimulus features such as colour, luminance, or semantic properties, for items they were to be aware of. In theory these parameters, known as the ‘attentional set’, make search more efficient by aiding the elimination of irrelevant distractor items. In terms of the current study, an attentional set created by a helpful instruction may prioritise objects cued by the people in the scenes. Aimola Davies, Waterman, White and Davies (2013) explored the potentially detrimental effects of an attentional set in a dynamic selective-looking task. Participants were told four squares and four diamonds would move across the screen, and that their task was to silently count how many times the four squares or diamonds (depending on trial number) bounced off the edge of the screen. In the critical trial, the authors added a critical stimulus to the array. This would be in the shape of objects being counted, but in the same colour as the distractor objects (e.g. if black squares were being counted the critical stimulus would be a white square). Participants were then asked if they had observed anything in addition to the usual four squares and four diamonds. Aimola Davies et al. (2013) reported that when participants had generated an attentional set primed for counting bounces by black squares they failed to notice a white square crossing the scene, even though it was visible for 5.5 s. However, if participants' attentional sets were primed to count white diamonds, they did notice the critical white square stimulus. These differentiations can be mapped on to the attentional sets that would be created by the different instructions used in the current studies. A helpful instruction would promote an attentional set that would first selectively process objects cued by the people in the scene. This is only beneficial in 50% of trials, for the remaining 50% of trials participants would then have to search the remaining objects after processing those cued by people in the scene causing delays in all aspects of performance, as is demonstrated in the results above. Conversely, when participants are told to ignore the people in the scene, the only attentional set they are primed to is the array of objects on the table.

While it is clear from the evidence discussed here that task instruction can change the way in which participants move their eyes, this evidence cannot account for why there is still generally poorer performance in instruction conditions that make people presence salient. Schneider, Nott and Dux (2014) may provide an explanation for this in their study examining Theory of Mind (ToM) in adult participants. As was discussed in Chapter One of this thesis, it is understood that ToM plays a key role in interpreting social cues provided by others. While the basic definition of ToM – that it is the means by which we can predict or explain other peoples' thoughts, beliefs and feelings – still stands, recent research has suggested people may in fact possess two separate ToM systems. Apperly and Butterfill (2009) suggest we have one system that operates implicitly (iToM), which is present from early life and used to monitor belief states of others. The second system develops later and operates in a controlled manner, allowing explicit ToM inferences (eToM). In their study, Schneider et al. (2014) ask participants to engage in one of three tasks, monitoring eye movements while they focused on one aspect of the Sally-Anne false-belief task. Eye movement data indicated that participants implicitly track the mental states of others (in this case, their beliefs about the locations of the hidden ball) even when this mental state monitoring is incongruent to the task they have been assigned. Perhaps most importantly, the authors reported that their participants were unaware of any such belief monitoring occurring. Schneider et al. (2014) state this is evidence in support of an implicit ToM system that operates regardless of top-down task settings. If this is indeed the case, it is possible that any instruction that makes the presence of people in the scene salient, regardless of whether they are told that these people are helpful or unhelpful to the task, will engage their iToM systems track the focus of the peoples' attention without them even being aware of it. Schneider et al. (2014) noted that this processing does draw on executive functioning resources, suggesting that using this iToM system may increase to some extent the cognitive load experienced by participants when performing the task in a people-salient instruction condition. This could explain why in some measures, any people-salient instruction – helpful or unhelpful – results in performance worse than that seen in the no instruction condition, with the only question remaining to what extent performance deteriorates depending on if participants believe the cues are helpful or unhelpful. This could also account for the interactions seen in the response time and error rate data, where having any instruction that made people presence salient affected performance in people present versus people absent trials.

It is clear that the evidence supports a hypothesis of task instruction changing the way in which observers view a scene. The overt gaze-seeking data also clearly demonstrates participants' fixations on the people in the scene change depending on what they are told about the purpose of people presence within the scene, which permits confidence in the assumption that the participants have believed the experimental manipulation. The results show that an instruction suggesting people presence may be helpful in finding the target object actually causes deterioration in performance in the task with slower first fixations on the target, slower overall response times and less efficient searches. Alongside previously discussed work by Yarbus (1967), more recent research such as that by Itier et al. (2007) and Fletcher-Watson et al. (2008) demonstrates differences in observer eye movements – even when they are shown the same stimuli – when they are given different task instructions. It is possible that the instructions, particularly in the helpful instruction condition, generate an attentional set that causes them to prioritise objects cued by the people in the scene. Since these cues are spatially-uninformative this type of attentional set will only benefit participants on half of trials. In the remaining trials they have to process these gazed-at objects before continuing search through the rest of the array, thus resulting in slower and less efficient searches. It is also possible that an implicit Theory of Mind system causes increased cognitive load in any people-salient instruction condition, which might explain the general deterioration in performance seen when either a helpful or unhelpful instruction is given. While this analysis provides strong quantitative evidence of the effects of instruction when cues are spatially uninformative, they do not address whether these effects persist if cues are actually helpful or unhelpful in locating the target object. To answer that question, the reliability of gaze cue senders would need to be manipulated so that they accurately predict target location more often than not (or vice versa in an unhelpful condition). Exploring this issue is the focus of the remaining experimental chapter, which examine how cues that are spatially informative are processed by the observers in each of the three instruction conditions.

Chapter Eight

Manipulation of gaze cue reliability in a visual search task featuring simultaneously-presented multiple gaze cues

Introduction

The previous chapters have established the realistic Posner-type paradigm with both single and multiple gaze cues, and have explored how observers respond to these cues in a visual search task. These chapters have also introduced different types of instruction to the paradigm, which are designed to manipulate the observers' perception of the purpose of people within the scene. Until this point, all of these studies have used non-predictive gaze cues, which mean that they are equally likely to cue the target as they are to cue the distractor, thus offer no real advantage for gaze-seeking and –following behaviours. However, using these types of cues makes it difficult to determine how observers are using the gaze information provided to them, particularly when they are presented with a conflicting gaze cue. This chapter aims to explore whether participants' eye movement behaviour changes when the cues they are given become spatially informative, and may reliably cue the target object they are searching for.

How we assess the reliability of a cue presents an issue when navigating our environment. Jacobs (2002) highlights the many ways in which this can be problematic in real world visual environments. He states that because all visual cues are ambiguous, cue reliability can depend largely on the context in which it occurs. This applies for all types of visual cues, not just gaze cues, but the same principles apply for observers assessing reliability in the studies described within this chapter. Jacobs (2002) explains that to determine if a cue is reliable, the information it provides must be assessed in relation to the information given by other cues in the environment. According to Jacobs (2002) observers assess cue reliability by making correlations among cues; cues are rarely congruent with each other accidentally – instead this occurs because of some underlying feature of the environment. The more cues that are in agreement, the more reliable they are. In terms of the stimuli used previously, by Jacobs' (2002) reasoning, the congruent and incongruent gaze cue conditions would be deemed most reliable, because these exhibit the highest number of cues in agreement. The conflicting gaze cue condition makes it more difficult for observers to assess reliability in this way.

It has been documented that spatially informative cues produce quite different responses from more commonly used spatially uninformative cues. The previous chapters within this thesis have used spatially uninformative cues in the sense that these cues predict target location in 50% of trials, and are therefore no more or less likely to predict the location of the target object on one trial compared to another. In reality, theses cues are not truly uninformative, because if an observer followed the gaze cues consistently they would still be directed toward the target in half of trials, compared to only one-fifteenth of trials if no cues were present (because the target object is one within an array of 15). However, to truly examine the effect of spatially informative cues, the proportion of trials in which the target is cued needs to be either much higher or much lower, to be either reliable or unreliable. Martín-Arévalo, Kingstone and Lupiáñez (2013) discuss how providing a centrally-presented spatially informative cue results in long-lasting benefits to task performance, most often measured in the form of faster response times. Uninformative cues, however, produce different responses. When observers are presented with short stimulus onset asynchronies (SOAs) their performance is facilitated, but a long SOA induces longer response times and more instances of inhibition of return (Funes, Lupiáñez & Milliken, 2007).

This issue was explored Driver et al.'s (1999) research. This research has been discussed previously, but here discussion will focus on the third experiment described in their paper. In this experiment Driver et al. (1999) presented participants with a spatially informative cue that was consistently unreliable: it cued the side of the screen away from the target four times more frequently that it cued the side on which the target would appear. Rather than leaving it to participants to infer this reliability through trial and error, Driver et al. (1999) told their participants that the cue provided would be consistently unreliable in predicting target location, and they were reminded of this at the beginning of every block of trials. At short SOAs (300 ms) participants responded fastest to the target when it appeared on the side the face cued, even though participants knew this would happen very rarely. However, when the SOA was increased to 700 ms participants' expectations of target location changed, and they were faster to find targets on the side the face gazed away from. Driver et al. (1999) used this evidence to support a hypothesis of reflexive orienting because participants response times did not improve at the gazed-away-from side until a 700 ms SOA was used. Regardless, this also provides evidence that observers can apply the reliability of a gaze cue to modify their search. Considering Driver et al.'s (1999) evidence in terms of the current studies, it could be predicted than when cues presented are reliable participants should perform better in the search task than when cues are unreliable.

Friesen, Ristic and Kingstone (2003) followed a similar procedure to explore volitional orienting to eye gaze cues. They used counter-predictive cues to further test out the different levels of reflexive orienting found by Driver et al. (1999) at their two SOA points. Rather than just presenting the spatially informative gaze cue at two SOAs, Friesen et al. (2003) used four SOAs ranging from 105 ms to 1800 ms. Trials were in one of four conditions: the target would appear in the predicted location (i.e. the opposite side to where the face gazed toward), the target appeared in the gazed-at location, a not predicted condition where the target appeared somewhere that was not predicted or cued, or a no cue condition where the face gazed straight ahead. Friesen et al. (2003) also found that it was only at longer SOAs that participants could divert their gaze from reflexive orienting to volitional orienting, with faster response times only occurring for the predicted location (opposite to the gazed-at location) in these longer SOA trials.

The evidence from Driver et al. (1999) and Friesen et al. (2003) demonstrates that varying the reliability of gaze cues does change observer eye movement behaviour. Their evidence comes with the caveat that these effects only become apparent when SOAs are a minimum of 700 ms. However, as previous chapters have discussed, these studies use only a single gaze cue provided by a schematic line-drawing of a face. Therefore, it is difficult to be confident when applying these findings to real world gaze behaviour. The current study aims to address this by presenting observers with realistic stimuli that are more representative of the sorts of gaze cues experienced in the real world. Unlike Driver et al.'s (1999) study, participants will be unaware of the reliability manipulation, which allows investigation into how participants discern reliability for themselves, or if any reliability effects emerge at all. The method, results, and discussion below discuss three studies: the reliability manipulation has been repeated across all three instruction conditions. These are the same as in previous chapters – participants are either given no instruction, a helpful instruction, or an unhelpful instruction. Results are presented measure-by-measure, comparing performance across reliability condition rather than gaze cue condition.

Method

Participants

A total of 22 people (9 male) were recruited for participation in the no instruction condition, 20 people (7 male) were recruited for participation in the helpful instruction condition, and a further 20 people (7 male) were recruited for participation in the unhelpful instruction condition. All had normal or corrected vision and were naïve to the purposes of the study. Level one and two undergraduate students received course credits for participation; anyone not eligible for course credit was paid £2. None of the participants had taken part in any of the previous studies.

Materials

The materials used in this experiment were similar to those described in Chapters Five and Six. However, in order to accurately manipulate the reliability of each person in the scene, only conflicting gaze cue condition scenes were used (Figure 69). Using these scenes means that in both versions of the experiment only one of the two people is consistently reliable; scenes where both people look toward the same object would interfere with the proportion of trials in which Person A was reliable compared to Person B. In both versions of the experiment one person cues the target object with 70% reliability, and the second person cues the target only in 30% of trials. There are two versions of the experiment to counterbalance which individual pictured in Figure 69 is the reliable gaze cue sender.

To recap the basic composition of scenes, each scene featured one of the ten sets of 15 everyday items arranged on a table top. Within each scene one item was designated the target and another designated the distracter. These items were always on the opposite sides of scene centre, preventing any central bias (as discussed in Tatler, 2007). The target was equally likely to appear on the left or right side of the table.

Figure 69. Examples of scenes used for one arrangement of objects on the table top. In this case the target item was the Filofax, and the distractor was the pair of earmuffs.

As in previous experiments there were a large number of factors which required counterbalancing within this study: the position of the target and distractor; the number of times the target/distractor appeared on the left or right for each participant; and the position of each individual within the scene. There were two versions of the experiment created. In version one, one person would reliably cue the target on 70% of trials. In the second version, they would only cue the target on 30% of trials. Since every people present trial shows a conflicting gaze cue, the comparison for this chapter is across reliability condition, rather than gaze cue condition. Three conditions were used in this experiment: *people absent*, *reliable person looking at target*, and *unreliable person looking at target.*

In Figure 69 panels A and B would have both been shown to participants in both versions of the experiment. The difference between these trials is the target object they were asked to search for. For example, if Person 1 (with the top-knot bun) was the reliable gaze cue sender, the target object paired with Panel A would be 'earmuffs', and the target paired with Panel B would be 'Filofax'. The 30% of trials in which the reliable person did not cue the target were randomly selected. As in previous experiments, there were people absent scenes for each set of objects, each of which was shown twice.

Eye Tracking

Participants' eye movements were tracked using an SR Research EyeLink 1000 eye tracker with a sampling rate of 1000 Hz, using pupil tracking and corneal reflection. The tracker was desk-mounted, sitting below the computer monitor and used to track a participant's dominant eye. The participant's head was kept stable throughout the experiments using an adjustable chin and forehead rest. Stimuli were presented on a 19-inch CRT computer monitor with a resolution of 1024 x 768 pixels. The Experiment Builder software developed by SR Research was used to run the experiment. Calibrations performed using the EyeLink 1000 were accepted if the average spatial error was less than 0.5 degrees and the maximum error was less than 1 degree over the 9 calibration points.

Procedure

Procedure was the same as that described in Chapter Five. To reiterate: a single-point calibration check was performed before each trial began. The name of the target object for the trial was presented on a grey-scale background for 500 ms. Most target object names were high frequency words, but to control for variation in participant vocabulary, a 500 ms presentation time was used following mean naming time of written words across low to high frequencies established as 546 ms by Schilling et al. (1983). This was followed by the presentation of a blank screen for a further 500 ms. In experiments where participants freely view scenes, it is common practise to include the presentation of a white noise mask for 500 ms following scene presentation to prevent interference between trials (e.g. Tatler, 2007; Tatler & Vincent, 2008; Tatler & Vincent, 2009). In the current study a 500 ms blank screen was shown prior to scene presentation to prevent interference from any residual word processing following presentation of the target word. After this blank screen presentation, the visual search scene appeared. To indicate they had found the target, participants were asked to press either of the trigger buttons on a Microsoft Sidewinder gamepad – whichever they found most comfortable to use. Scene presentation ended with the button press or after 10 s had elapsed with no response. Each participant saw a total of 100 scenes: 20 people absent scenes and 80 person present scenes. This means that every participant saw every image twice, once searching for the target object and once searching for the distractor.

In the no instruction condition, participants were given no information regarding the presence or absence of a person in the scenes. They were given a brief description of what would happen in each trial and simply asked to find the target object as quickly as possible. In the unhelpful instruction condition, participants were told to ignore the presence or absence of people in the scene, similar to previous Posner-type tasks, with the instruction: "*Some of the scenes will have people in them, but please just ignore them. I'm using the same images over several experiments, but in this experiment the person isn't relevant; I'm only interested in how you search for the target object in the scene.*" Conversely, in the helpful instruction condition participants were told: "*Some of the scenes will have people in them. One of these people might be looking at the target, so they may help you find it faster.*" This instruction does not tell participants that they must look at either person in the scene; it simply provides them with more information about the context of the scene. It should also be noted that participants are explicitly told only one of the cue-givers might be helpful

Data for the study using no instruction was collected in part by a Research Assistant, Gemma Mackintosh, under my supervision. Prior to collecting data on her own, Gemma was trained to use the EyeLink 1000 eye tracker and the calibration procedure used for each participant. We practised instructions given to participants prior to the experiment, and the debriefing given after they had completed the experiment. Gemma observed me collect data for 4 of the 22 participants, then ran one test participant under my supervision whose data was not included in the final analyses. Gemma collected data from four participants under my direct observation, then from another two participants with me in an adjoining room should she have any problems with the equipment. No issues occurred. She collected data from the remaining 12 participants alone.

Data Analysis

Data were analysed for these experiments following the same procedure as described in Chapter Five using the R statistical analysis environment (R Development Core Team, 2011). To reiterate the salient points: lmer() functions return *z*- and estimated *p*-values for logistic models and *t*-values for linear models, within which we consider any effects for which the *t*-value is greater than two as reflecting a significant effect (as in Kleigl et al., 2012). In all models the random factors of participant and scene were included, and where possible the maximal model was used in which intercepts and slopes for the fixed effect of gaze cue condition was allowed to vary over both of the random factors (see Bates et al., 2014). Maximal models often fail to converge when large amounts of data are unavailable so random effects structures were simplified when necessary. In the analyses below the most complicated random effects structure that converged is reported.

Two different stages of search are reported in the analyses: search initiation (first saccade latency, first saccade direction, first saccade end point accuracy), and scene scanning (time to first fixate target, scan path ratio). Overall search behaviour is analysed in terms of the response times of participants to press the button, terminating search. Behaviour with respect to the individual pictured in the scenes (number of looks at the person) is also considered. The following results compare performance in each of these measures across reliability conditions rather than in gaze cue conditions used in the previous chapters. Performance in each instruction condition is discussed within the sub-heading for each measure.

Results

First Saccade Latency

The first saccade latency data presented a small number of very short latencies, which most likely were the results of pre-emptive eye movements beginning before the appearance of the scene. Very short latencies were defined as being less than 100 ms, and these were removed from the dataset after they had been identified. A total of 266 trials in the no instruction condition featured a very short latency (12.66% of the total number of trials). In the helpful instruction condition 269 trials featured a very short latency (13.46% of total trials), and in the unhelpful instruction condition 287 trials featured a very short latency (14.36%). Figure 70 below presents these data with the very short latencies excluded, prior to undergoing logarithmic transformation to generate a normal distribution.

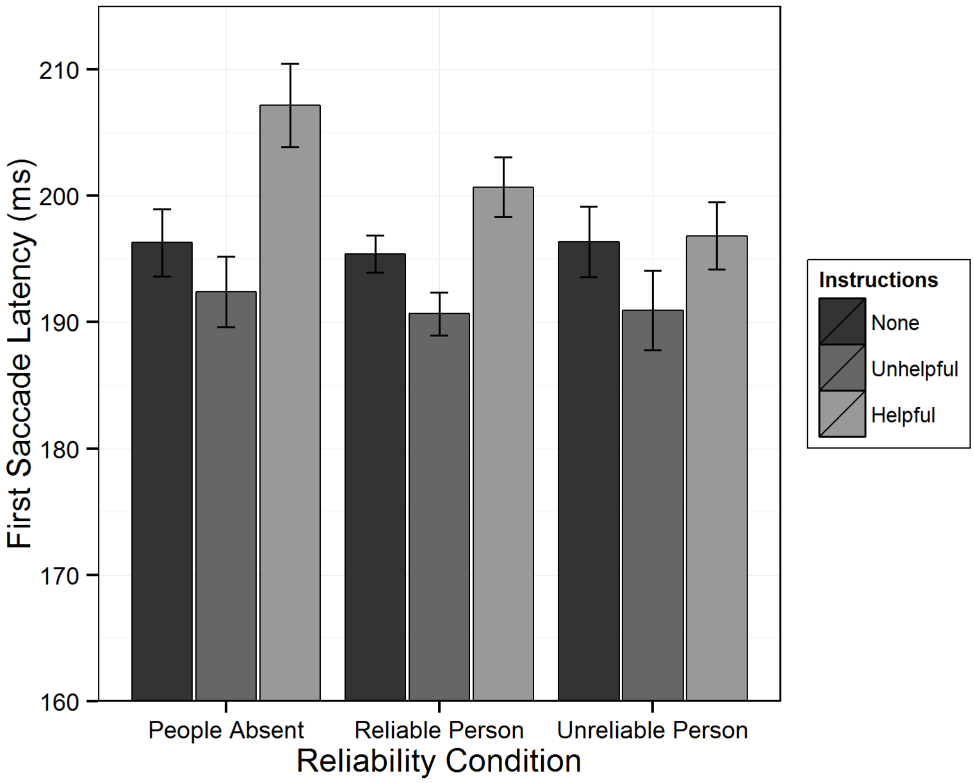

Figure 70. First saccade latency (ms) across all three reliability conditions in each instruction study, with error bars displaying standard error across all data samples. The dark grey bar represents the no instruction study, medium grey bar represents the false study condition and the light grey bar represents the true instruction study.

As might be expected from Figure 70, in the no instruction condition neither of the people present reliability conditions showed any change in first saccade latency when compared to the people absent condition (*t*s < 0.5). Similarly, there was no difference in first saccade latency between the reliable person and unreliable person looking at target conditions ($t < 0.5$).

This pattern of results persisted in the unhelpful instruction condition with no changes in first saccade latency evident when the people present reliability conditions were compared to the people absent condition (*t*s < 1). First saccade latencies in the reliable and unreliable person looking at target conditions were even more similar ($t < 0.5$).

However, in the helpful instruction condition people presence seemed to benefit performance. Compared to the people absent reliability condition, the unreliable person, β = -0.016, *SE* = 0.007, *t* = -2.16, and reliable person conditions, β = -0.014, *SE* = 0.006, *t* = -2.22, resulted in faster first saccade latencies. However, there was no difference in first saccade latency between the people present reliability conditions ($t < 0.5$).

First Saccade Direction

The proportion of first saccades directed toward the target showed greater variation across reliability conditions, as can be seen in Figure 71.

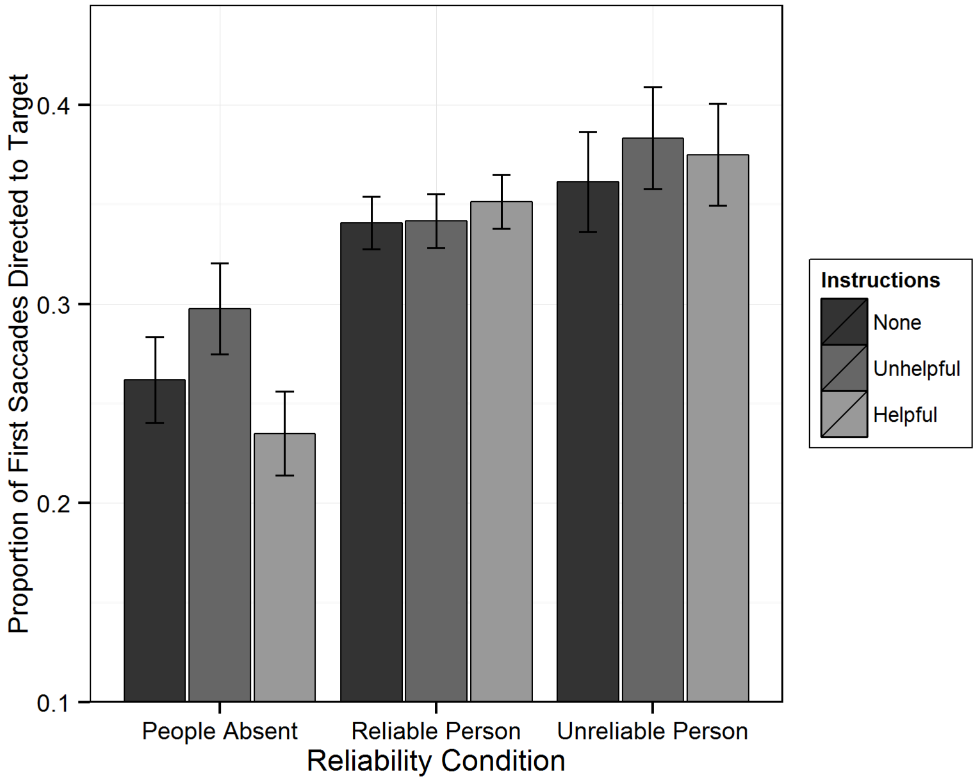

Figure 71. The proportion of first saccades directed toward the target across all three reliability conditions in each instruction study, with error bars displaying standard error across all data samples.

In the no instruction condition both the reliable person, $\beta = 0.078$, $SE = 0.031$, $t = 2.472$, and unreliable person looking at target conditions, $\beta = 0.101$, $SE = 0.036$, $t = 2.765$, produced significantly higher proportions of first saccades directed toward the target compared to the people absent condition. However, there was no difference in the proportion of first saccades directed toward the target between the people present reliability condition ($t < 1$).

While there was some benefit of people presence in the unhelpful instruction condition, it did not have as strong an effect. Compared to the people absent condition, both the reliable person, $\beta = 0.044$, $SE = 0.029$, $t = 1.478$, and unreliable person conditions, $\beta = 0.085$, $SE = 0.043$, $t = 1.978$, resulted in higher proportions of first saccades directed toward the target, but neither reached significance (though the unreliable person condition is approaching significance). While the proportion of first saccades directed toward the target in the unreliable person looking at target condition was somewhat higher than that in the reliable person condition, $\beta = 0.041$, $SE = 0.033$, $t = 1.240$, this was not significant.

The results in the helpful instruction condition were more similar to those in the no instruction condition. The reliable person, $\beta = 0.116$, $SE = 0.030$, $t = 3.779$, and unreliable person, $\beta = 0.140$, $SE = 0.034$, $t = 4.032$, conditions both resulted in a greater proportion of first saccades being directed toward the target than in the people absent condition. Performance in the reliable and unreliable conditions was similar, with no difference in the proportion of first saccades directed toward the target ($t < 1$).

End Point Accuracy of the First Saccade

Figure 72 presents the end point accuracy data across all three reliability conditions in each instruction study.

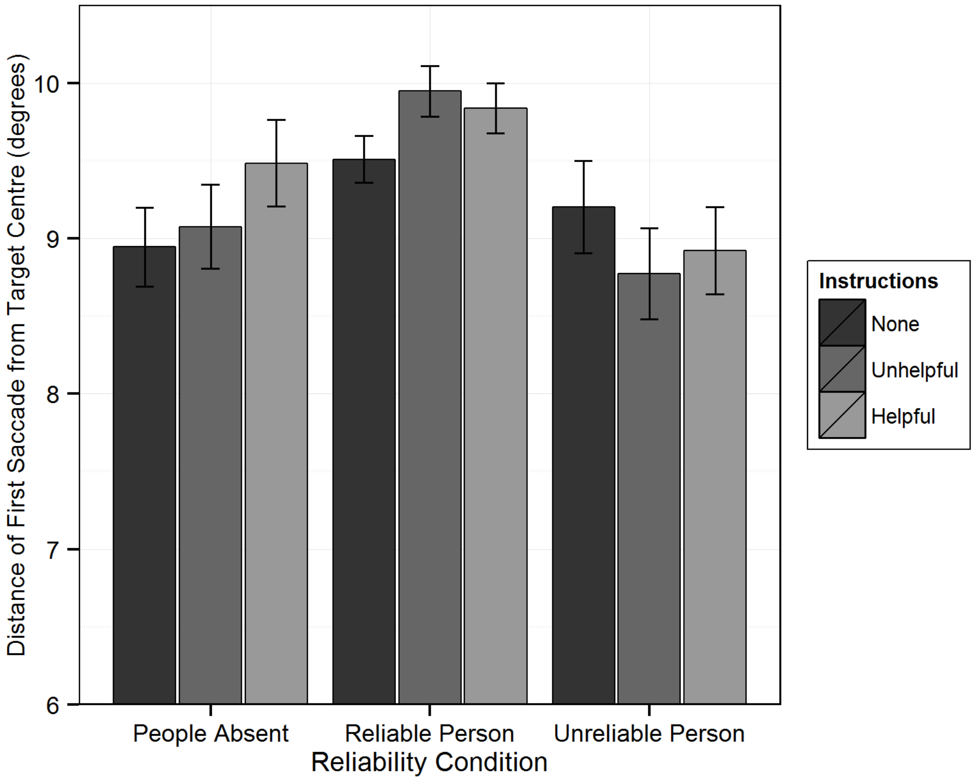

Figure 72. The distance of the landing point of the first saccade from the centre of the target boundary box (in degrees of visual angle) as a measure of end point accuracy across three reliability conditions in each instruction study. Error bars show standard error across all data samples.

Analysis of the no instruction condition showed that the end point accuracy found in the people absent condition was similar to that seen when the unreliable person cued the target ($t < 1$). However, when a reliable person cued the target the end point accuracy deteriorated, becoming less accurate than the people absent reliability condition, $\beta = 0.580$, $SE = 0.301$, $t = 1.925$, though this was only approaching significance. While there was deterioration evident numerically in the reliable person condition compared to the unreliable person condition, $\beta = -0.378$, $SE = 0.306$, $t = -1.235$, this was not statistically significant.

Results were similar in the unhelpful instruction condition. When the unreliable person cued the target end point accuracy was similar to that seen in the people absent reliability condition ($t < 1$), however when the reliable person cued the target end point accuracy deteriorated significantly, $\beta = 0.873$, $SE = 0.302$, $t = 2.886$. When the people present reliability conditions were compared it was found that the reliable person condition was also significantly less accurate than the unreliable person condition, $\beta = -1.174$, $SE = 0.314$, $t = -3.741$.

In the helpful instruction condition end point accuracy was similar across the three reliability conditions. The reliable person looking at target condition produced less accurate first saccades than the people absent condition, $\beta = 0.354$, $SE = 0.316$, $t = 1.120$, whereas the unreliable person looking at target resulted in more accurate first saccades than the people absent condition, $\beta = -0.561$, $SE = 0.378$, $t = -1.484$. When the people present reliability conditions were compared it was found that the unreliable person condition produced significantly more accurate first saccades that the reliable person condition, $\beta = -0.915$, $SE = 0.312$, $t = -2.934$.

Time to First Fixation on Target

The time taken to first fixate on the target object was the first measure of the scene scanning phase of search. The data are presented in Figure 73 untransformed: data required logarithmic transformation for analysis to satisfy model assumptions.

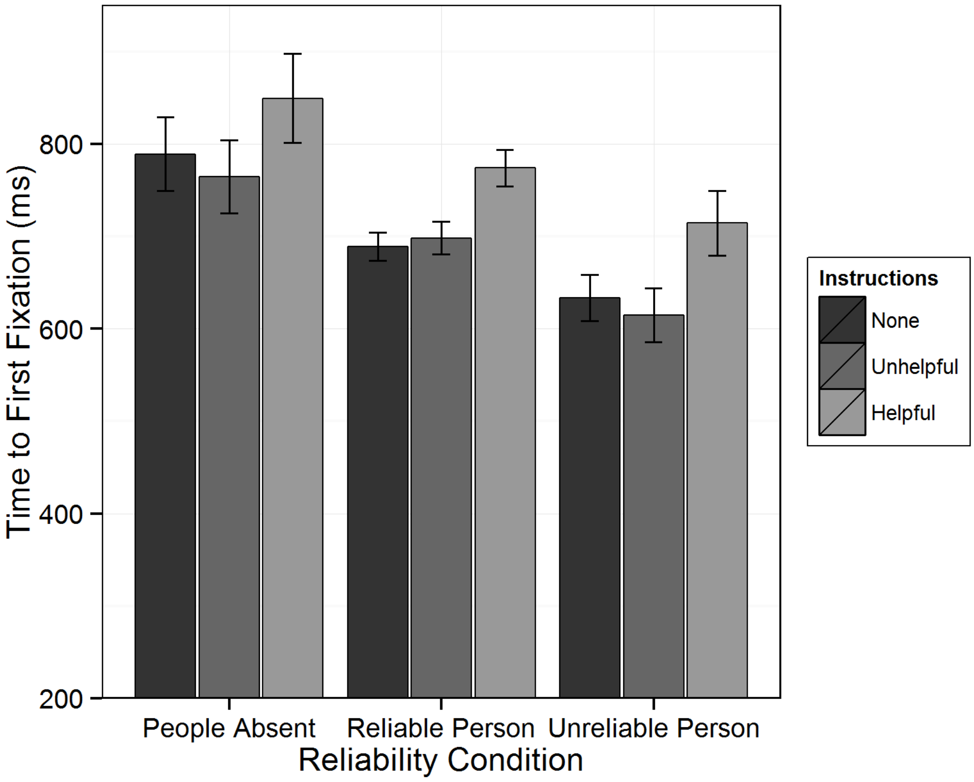

Figure 73. The time to first fixation on the target (ms) from scene presentation across three reliability conditions in each instruction study. Error bars show standard error across all data samples.

Analyses of the no instruction condition showed that both the reliable person, β = -0.041, *SE* = 0.022, *t* = -1.83, and unreliable person reliability conditions, β = -0.076, *SE* = 0.026, *t* = -2.84, resulted in faster first fixations on the target than the people absent condition, though only the latter was significantly different. When the reliable and unreliable person looking at target conditions were compared it was found that the first fixations on the target in the unreliable person condition were significantly faster than those in the reliable person condition, β = -0.034, *SE* = 0.016, *t* = -2.11.

In the unhelpful instruction condition the time taken to first fixate on the target is relatively similar between the people absent and reliable person conditions (*t* < 0.5). However, when the unreliable person condition was compared to the people absent condition it was found that there was a significant decrease in the time taken to first fixate on the target, $\beta = -0.071$, $SE = 0.026$, $t = -2.66$. The unreliable person condition also produced faster first fixations on the target than the reliable person condition, $\beta = -0.062$, $SE = 0.018$, $t = -3.43$.

When given a helpful instruction, there were minimal differences apparent in the time taken to first fixate on the target across the three reliability conditions. While the unreliable person looking at target condition resulted in somewhat faster first fixations on the target than the people absent condition, $\beta = -0.056$, $SE = 0.032$, $t = -1.73$, this is not significant; nor is the slight decrease in time taken to first fixate on the target in the reliable person condition ($t < 1$). Further analysis did show a significant improvement in time taken to first fixate on the target in the unreliable person condition as compared to the reliable person condition, $\beta = -0.041$, $SE = 0.019$, $t = -2.15$.

Response Time

As in the time until first fixation measure, the response time measure also required logarithmic transformation prior to analysis. However, data are presented untransformed in Figure 74 below.

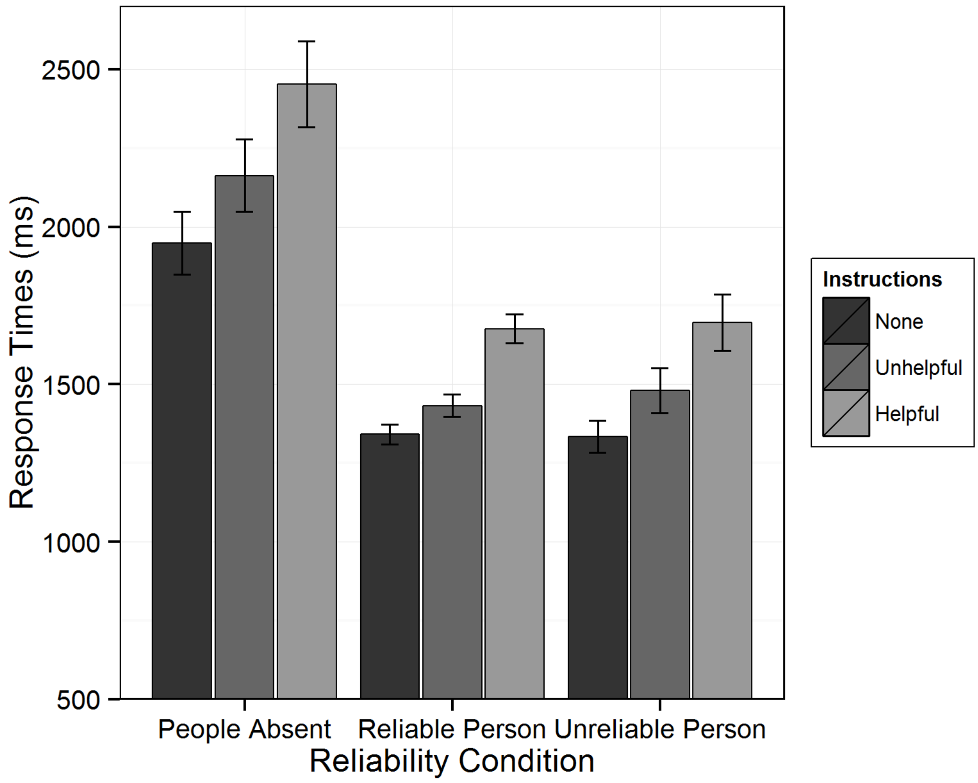

Figure 74. Response times (ms) to button press indicating successful search for the target across three reliability conditions. Error bars show standard error across all data samples.

Analysis showed that in the no instruction condition both the reliable person condition, $\beta = -0.141$, $SE = 0.016$, $t = -8.73$, and the unreliable person condition, $\beta = -0.134$, $SE = 0.018$, $t = -7.26$, produced significantly faster response times to find the target than the people absent condition. When the two people present reliability conditions were compared, however, there were no differences in response time ($t < 1$).

There were similar differences in response time observed in the unhelpful instruction condition. The reliable person, $\beta = -0.151$, $SE = 0.017$, $t = -8.55$, and unreliable person conditions, $\beta = -0.150$, $SE = 0.020$, $t = -7.49$, both resulted in faster response times than the people absent condition. However, there was no difference in response times observed between the reliable person and unreliable person conditions ($t < 0.5$).

The benefit of people presence continues in the helpful instruction condition. When compared to the people absent condition, both the reliable person, $\beta = -0.135$, $SE = 0.016$, $t = -8.16$, and unreliable person conditions, $\beta = -0.142$, $SE = 0.019$, $t = -7.26$, resulted in significantly faster response times. Further analysis found no differences in response times of the two people present reliability conditions ($t < 0.5$).

Scan Path Ratio

The no instruction condition scan path ratio, which provides a measure of search efficiency, also showed large differences in performance across reliability conditions. Compared to the people absent condition, both the reliable person, $\beta = -0.894$, $SE = 0.187$, $t = -4.781$, and unreliable person conditions, $\beta = -0.877$, $SE = 0.221$, $t = -3.963$, resulted in significantly more efficient searches with scan path ratios closer to 1. However, there was no difference in search efficiency between the people present reliability conditions ($t < 1$). This is shown in Figure 75.

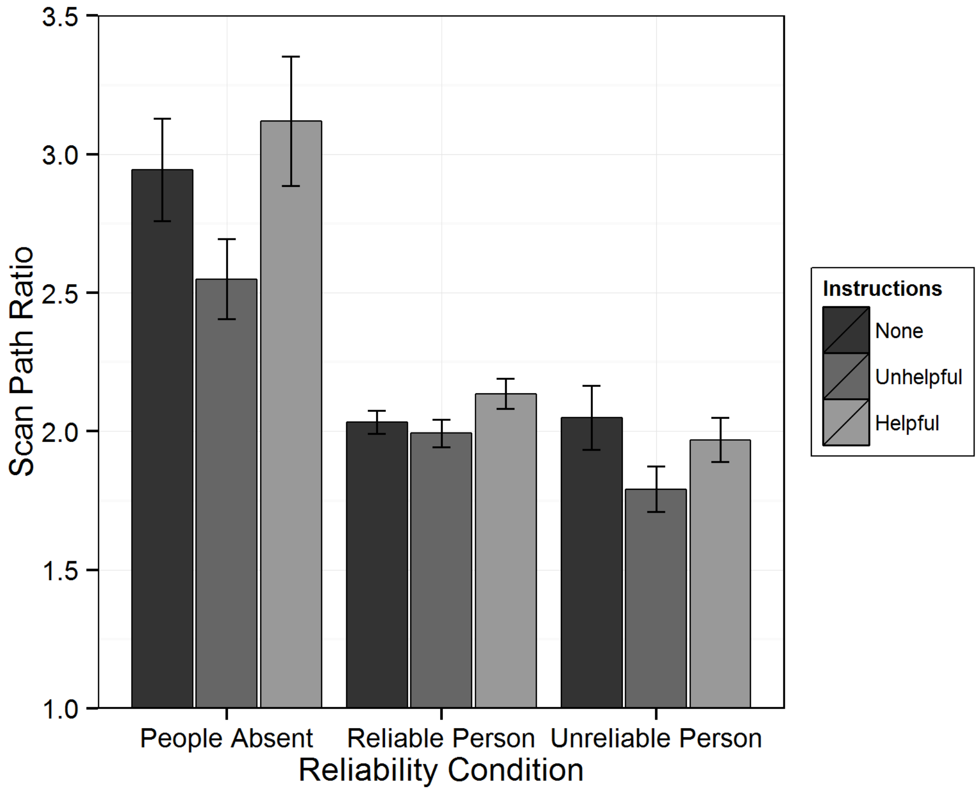

Figure 75. The scan path ratio across three reliability conditions in each instruction study. Error bars show the standard error across all data samples.

When participants were given an unhelpful instruction, the benefit of people presence to performance persisted. Both the reliable person looking at target condition, β = -0.557, *SE* = 0.165, *t* = -3.362, and the unreliable person condition, β = -0.760, *SE* = 0.163, *t* = -4.650, resulted in a more efficient search than that achieved in the people absent condition. While search was somewhat more efficient in the unreliable person condition compared to the reliable person condition, β = -0.203, *SE* = 0.117, *t* = -1.731, this was not significant.

Analyses in the helpful instruction condition showed that people presence improved search efficiency. The reliable person, β = -0.974, *SE* = 0.206, *t* = -4.721, and the unreliable person conditions, β = -1.139, *SE* = 0.259, *t* = -4.416, produced significantly more efficient searches than the people absent condition. When the people present reliability conditions were compared there was some difference evident: the unreliable person condition had a slightly lower scan path ratio than the scan path ratio in the reliable person condition, $\beta = -0.164$, $SE = 0.140$, $t = -1.173$, but this was not significant.

Error Rate

Error rate data – that is, measures of false-positive responses made by participants during search – is presented in Figure 76 below.

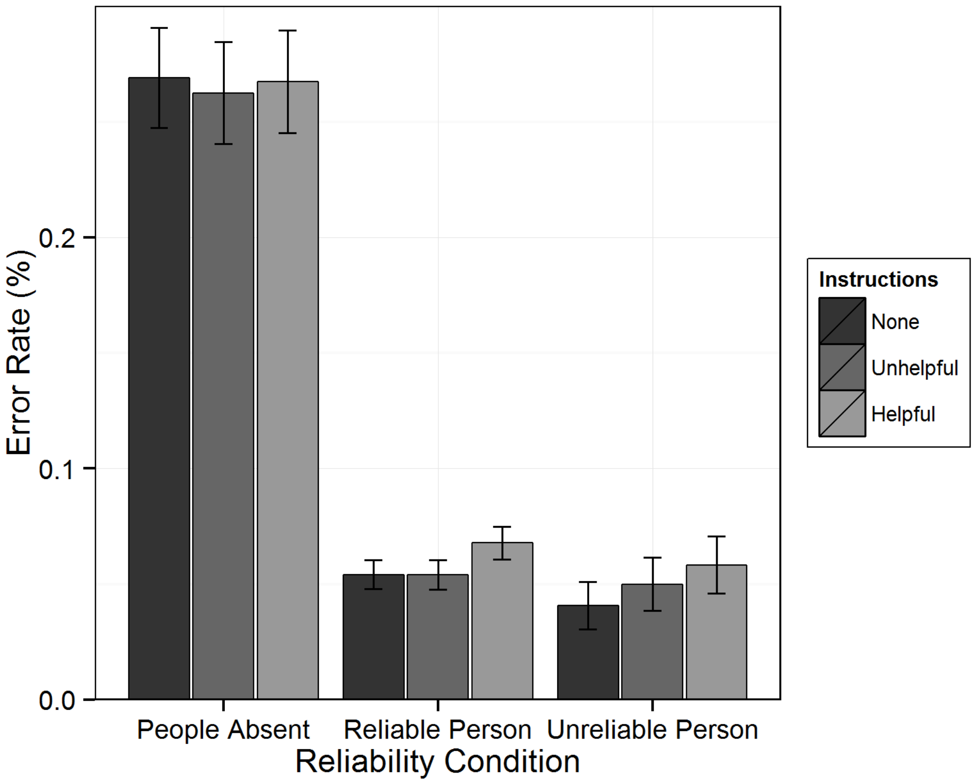

Figure 76. The error rate across three reliability conditions in each instruction study, shown as a percentage of trials in which participants made a false-positive response. Error bars show the standard error across all data samples.

The rate of false-positive responses in the no instruction condition was reduced considerably once people were present in the scene in the no instruction condition.

Compared to the people absent reliability condition, both the unreliable person, β = -0.229, SE = 0.019, t = -11.648, and the reliable person looking at target conditions, β = -0.214, SE = 0.015, t = -13.909, resulted in significantly less errors. Further comparison between people present reliability conditions showed no significant differences in the error rate ($t < 1$).

The benefits of people presence persisted in the unhelpful instruction condition. There were significantly lower error rates in the reliable person, β = -0.208, SE = 0.017, t = -11.659, and unreliable person conditions, β = -0.212, SE = 0.021, t = -9.910, than in the people absent condition. However, there was very little difference in error rates between the two people present reliability conditions ($t <$ 0.5).

The considerable benefit of people presence was also evident in the helpful instruction condition. Both the reliable person, β = -0.199, SE = 0.016, t = -11.963, and unreliable person conditions, β = -0.209, SE = 0.021, t = -9.928, produced significantly less false-positive responses than the people absent condition. When error rate was compared between people present reliability conditions no difference was found ($t < 1$).

Overt Gaze-Seeking

As a reminder, overt gaze-seeking was examined by using only ‘correct’ trials – trials in which no false-positive response occurred, and an instance of overt gaze-seeking was classified as a fixation that fell within the boundary box around either person’s face. If participants were given no instruction regarding people presence, overt gaze-seeking was very limit. Only 5.85% of 1594 valid trials featured a fixation on either person's face (93 trials). Statistical comparison of the people present reliability conditions showed that there was slightly more overt gaze-seeking in the unreliable person condition than the reliable person condition, $\beta = 0.014$, $SE = 0.011$, $t = 1.284$, but this was not significant.

When participants received an unhelpful instruction, fixations on either person in the scene were very infrequent. In 1513 valid trials only 4.43% featured a fixation on either person in the scene (67 trials). There was very little overt gaze-seeking in either people present reliability condition, and no statistical difference between them ($t < 0.5$).

There was more evidence of overt gaze-seeking when participants were given a helpful instruction: 13.35% of the 1492 valid trials featured fixations on either person's face (199 trials). There was clearly more overt gaze-seeking in the reliable person condition compared to the unreliable person condition, $\beta = -0.026$, $SE = 0.017$, $t = -1.509$, but this was not significant.

Discussion

The studies discussed in this chapter examine the effect of manipulating gaze cue sender reliability on eye movements in a visual search task. Rather than using the spatially uninformative cues documented in Chapters Two through Seven, this chapter takes the same paradigm but provides participants with one spatially informative cue; one person cues the target object on 70% of trials, and the second person cues the target in 30% of trials. This chapter documents the effects of the reliability manipulation in three different studies, each giving participants a different instruction about the purpose of people presence in the scene. These instructions are the same as discussed in previous chapters: participants were either told nothing about the people in the scene, to ignore the people in the scene because they were unhelpful, or that one person in the scene may be helpful in finding the target.

Generally the results suggest that the beneficial effects of people presence evidenced in earlier chapters persist when spatially informative cues are presented to participants, regardless of the type of instruction they are given concerning the purpose of people within the scene. The overt gaze-seeking measure permits confidence in the instruction manipulation as the percentage of trials featuring a fixation on either person's face more than doubles in the helpful instruction compared to the no instruction study, and triples the percentage seen in the unhelpful instruction study. When reliability effects do emerge, it seems that the unreliable person looking at the target improves performance more than when a reliable cue is given, which is unexpected. These effects are stronger in the scene scanning phase of search, but do begin to occur in search initiation.

That these findings also document the benefits of people presence in the scene to participant performance supports the conclusions of earlier chapters. While there does seem to be some evidence that participants believe the instruction they are given – demonstrated by the increase in overt gaze-seeking in the helpful instruction study compared to the unhelpful and no instruction studies – any overt gaze-seeking that occurs is still in a small percentage of trials, which suggest participants are seeking out gaze information covertly. As was discussed in Chapter Two, this type of gaze processing can be very helpful; Carrasco and McElree (2001) advocate that using covert attention allows an observer to process information more quickly. By covertly attending information at a cued location, this information is granted priority in later visual processing. Further benefits of people presence were discussed in Chapter Five, where the role of multiple gaze cues within search scenes was examined. According to Malcolm and Henderson (2009), gaze information provided by the people in the scene may help inform the observer's target template, adding more weight to the objects in the scene cued by either or both individuals so that subsequent eye movements are guided to these spatially-weighted areas. The studies documented within this chapter show that these effects persist when participants are presented only with conflicting gaze cues.

Effects of reliability were mixed across the performance measures analysed. Within the search initiation phase the effects of reliability were not consistent. There were only changes to first saccade latency in the helpful instruction study, where the people absent condition produced shorter latencies than both people present reliability conditions. However, the proportion of first saccades directed to the target were greater in both the reliable person and unreliable person looking at target conditions than the people absent condition when participants were given no instruction regarding people presence or a helpful instruction. Not all of these saccades were very accurate. In the unhelpful instruction study the people absent and unreliable person conditions produced more accurate first saccades than the reliable person condition, and the unreliable person condition also produced more accurate first saccades than the reliable person condition in the helpful instruction study. Generally speaking, for the search initiation phase in most cases the unreliable person condition elicited the best results, but this is not true for all measures.

These findings are contradictory to what previous research would predict. For example, in their study examining the effects of social status on observer eye movements, Foulsham, Cheng, Tracy, Henrich and Kingstone (2010) equated perceived status with expertise. Foulsham et al. (2010) showed participants four different video clips that each displayed three unacquainted individuals completing a decision-making task that required the individuals to rank a list of items in order of necessity for use in a survival situation. These clips had been peer-rated to assess the social status of each individual in the clip, and found that there were clear differences in social status rankings of each individual, with one individual having a much higher social status than the other two. Foulsham et al. (2010) found that eyes were preferentially fixated over any other part of the face, and far more than the torsos or other body parts, a finding consistent with previous literature (e.g. Birmingham et al., 2008; Fletcher-Watson et al., 2008; Gullburg & Holmqvist, 2002; Klin, Jones, Schultz, Volkmar & Cohen, 2002). The perceived social status was also found to be a strong predictor of where fixations would occur. Individuals in the video clips rated to have high social status – those who were perceived as leading the task or the group – were fixated more often and for longer than the other individuals in the clips. Foulsham et al. (2010) hypothesised that social status influenced fixations because someone with high social status in a group is perceived to have a greater level of expertise. Previous literature would support this hypothesis: Henrich and Gil-White (2001) describe how humans' evolutionary history has led to social learning through which we attribute 'prestige'; what we could refer to as expertise. In the early development of our species when survival was challenging, a preference for attending an individual who had a greater level of expertise, be it in a practical skill or applied knowledge, increased chances of survival. Therefore, attending a reliable source of information has an extended evolutionary history.

Considering Foulsham et al.'s (2010) findings in terms of the current study, it would be expected that the reliable person condition would elicit the best performance. In this reliability condition the person cues the target in 70% of trials, therefore displaying a greater degree of expertise in the task. This should be particularly pronounced in the helpful instruction study, where participants are explicitly told that one of the people in the scene may be helpful to task performance. However, differences between the reliable and unreliable person conditions do not occur very often in the search initiation phase, and when they do the unreliable person cuing the target produces better task performance than when the reliable person cues the target. This counter-intuitive effect of reliability continues in the scene scanning phase of search.

Reliability effects were slightly more consistent in the scene scanning phase, but still relatively infrequent. In the first measure of speed – the time taken to first fixate on the target object – the unreliable person condition produced the fastest first fixations on the target in all instruction studies, while the reliable person condition produced fixations at similar speeds as the people absent condition. Across all three studies the unreliable person condition elicited better performance than the reliable person condition, and in the no instruction and unhelpful instruction studies, it also resulted in better observer performance than in the people absent condition. Response times to indicate the target had been found were faster in both people present reliability conditions than the people absent condition in all three studies. These results suggest that an unreliable cue results in a faster first fixation on the target, but both reliable and unreliable cues facilitate overall search time. Search efficiency and error rate measures provide an insight into how efficient and successful search was across the three studies. For both of these measures, people presence improved performance. The reliable and unreliable person conditions produced more efficient searches with fewer error responses than the people absent condition across all three instruction types. There was no difference in performance between people present reliability conditions in either scan path ratio or error rate measures, suggesting that while people presence improved performance, the reliability of the cue did not have an effect.

Based on previous Posner-type tasks that have varied cue reliability, this would not be the expected outcome of the reliability manipulation in these studies. For example, Geng and Behrmann (2005) conducted an experiment that examined how spatial probability could operate as an attentional cue in a visual search task. They state that implicitly presented spatial probabilities – where participants are not explicitly told that reliability of cues will vary – are powerful determinants in visual processing. This definition of implicitly presented spatial probabilities would include the spatially informative cues in the no instruction and unhelpful instruction studies. However, few studies had examined these effects using standardised cues; something Geng and Behrmann (2005) wished to address. They presented participants with a search array comprised of a circle of orange letters (T and L) on a black background. Participants were asked to press different keys depending on wither the target letter T was rotated to the left or right. The probability of the target appearing in one location varied across different experimental conditions, ranging from 25% to 75% predictability. The authors found that highly probable locations for the target, indicated by reliable cuing, resulted in faster processing of targets that appeared within that area. Thus, greater reliability resulted in better performance.

However, in the current studies there was not a single instance of a reliable cue benefiting performance to a greater degree than an unreliable cue. Given the evidence described above, this finding is counter-intuitive. The scenes were available for a maximum of 10 s, certainly long enough for volitional gaze following to occur, based on the evidence described in Driver et al.'s (1999) and Friesen et al.'s (2003) research using different SOAs in visual search paradigms. Perhaps then, instead of providing evidence for utilising gaze cues, the results found in these studies suggest participants have – in some trials at least – disregarded the cues given to them.

It is possible that conflicting gaze cues present a unique processing difficulty. Cañadas and Lupiáñez (2012) conducted an inverted Stroop-task for gaze cues, where a face would appear to the left or right of a central fixation point, and cue either to the opposite side of the screen or to the same side of the screen. Participants had to determine the gaze direction of the face, and the different positioning and cue direction of the face created what the authors call 'spatial interference'. The spatial interference was strongest when the face appeared cued the same side of the screen as it appeared on (e.g. the face appeared on the right hand side of the screen and provided a gaze cue to the right). This condition resulted in longer response times and decreased accuracy in gaze direction determination. Cañadas and Lupiáñez's (2012) results would suggest that participants find discriminating gaze cues more difficult in the current experiments' stimuli presented in Panel A of Figure 69 than Panel B. Since half of the people present trials are comprised of the composition of gaze cues shown in Panel A (Figure 69), perhaps this difficulty in processing led to participants disregarding the gaze cues because it would have greater cost to successful task performance than benefit. This would not be true for all trials, as there is some evidence of overt gaze-seeking in all instruction studies, but may explain the very low percentage of trials in which this occurred, particularly in the no instruction and unhelpful instruction studies.

Another possible explanation for the lack of reliability effects is that the number of trials presented to participants in these studies was not sufficient for them to identify the reliable gaze cue sender. To properly assess the effects of predictability, Geng and Behrmann (2005) used a very large number of trials in their experiments. For example, they used 900 trials per participant over five different blocks in their first experiment, 1008 trials in eight blocks in their second experiment, and 600 trials in six blocks in their third experiment. Driver et al. (1999) used 540 trials in each of their experiments. Friesen and Kingstone (1998) asked participants to complete 1500 trials: 1000 on one day of testing and 500 on the second. Ricciardelli et al.'s (2002) experiments used 360 and 400 trials across their experiments. In comparison, the current studies asked participants to complete just 100 trials, only 80 of which featured people within the scene. It must be noted that the high number of trials in the studies stated here are due in part to a large number of experimental conditions (most commonly a range of SOAs) that required full counterbalancing, which was not an issue for the current studies. However, on reflection, it is possible that 80 trials was not enough exposure for participants to recognise that one person was more reliable than the other in the no instruction and unhelpful instruction studies, or to identify which person was the reliable gaze cue sender in any of the three studies. Furthermore, each of the studies reported here used simpler stimuli than were presented in the studies documented in this chapter. Given that this additional level of complexity already creates a greater level of cognitive demand, perhaps assessing reliability was one demand too many on an already-strained visual processing system. These assumptions cannot be confirmed or denied by the data currently available, but provide a starting point from which future research can stream-line the investigation of reliability effects in realistic visual search paradigms.

This chapter presents three studies in which the reliability of gaze cue senders is manipulated so that the cues they provide are spatially informative. The results show that the same benefits of people presence to performance in a visual search task are observed when these cues are spatially informative as when they are spatially uninformative. This also permits confidence in the assumption that manipulating gaze cue sender reliability does not dramatically change eye movement behaviour in observers in the task. There is clear evidential basis for investigating effects of reliability: humans' evolutionary development to prefer those who are perceived to have more expertise (Henrich & Gil-White, 2001) is supported by eye movement data presented by Foulsham et al. (2010). However, while there were some effects of reliability, these effects were counter-intuitive: performance was only ever significantly improved when an unreliable cue was presented. It is possible the spatial interference (described by Cañadas & Lupiáñez, 2012) increased the cognitive demands for gaze cue processing to a point where using gaze cues would be of greater cost than benefit to task performance. It is also possible that the number of trials used in these studies was not sufficient to give observers time to determine who the reliable gaze cue sender was, or that there was a reliable gaze cue sender in the no instruction and unhelpful instruction studies. Though it is not possible to determine this conclusively from the evidence available from the studies described here, it provides a direction for future research in this area. What must be concluded from the evidence here then is that people presence clearly benefits performance in a visual search task to a considerable degree, but the role of conflicting gaze cues in this process cannot yet be fully understood, and more research is required in this area.

Chapter Nine

General Discussion

Studying social attention is inherently challenging because social attention comprises of so many different behaviours and cognitive processes. To look at another person and understand what they are thinking, to determine their beliefs about their environment, and to predict their future behaviour requires input from a whole host of cognitive processes working together. Studying gaze is an ideal route to understanding social attention because social gaze is key to almost all social attention behaviours. As highlighted by Kobayashi and Koshima (1997) the structure of the eyes with the pupil and iris easily discernable from the white sclera makes them ideal non-linguistic communicators. Furthermore, the nature of foveated vision (see Land & Tatler, 2009) – that to process a stimulus, we point our eyes at it – means that eyes function as 'pointers' to whatever we are attending. It is these two attributes that make gaze the ideal vehicle through which our understanding of social attention can be furthered. Gaze direction provides a useful clue as to what someone is attending, and is also a useful tool for studying how people attend to another person's gaze cues.

The paradigm developed within this thesis to study social gaze was built on a foundation of established work using Posner-type tasks. Friesen and Kingstone (1998) were the first to adopt Posner's (1980) paradigm, which was originally designed to investigate low level aspects of visual attention, and to combine it with social cues. Their aim was to determine whether the reflexive gaze shifts by infants in response to a caregiver's gaze cue could be replicated in adults. By using the simple but powerful manipulation of a centrally presented face providing a gaze cue that was either congruent or incongruent with target location, Friesen and Kingstone (1998) were able to show that observers' gaze would reflexively orient to the peripheral location cued by the face, even when they knew this cue was spatially uninformative and unhelpful in completing the task. This established the foundation for the model of reflexive orienting in response to gaze cues; a model that quickly gained further support (for examples, see Driver et al., 1999; Friesen et al., 2003a, 2003b; Hunt & Kingstone, 2003; Ricciardelli et al., 2002).

However, the ecological validity of using deconstructionist laboratory-based paradigms for investigating a complex social behaviour was questioned when real world research into social attention began to evidence social behaviours that contradicted the responses predicted by laboratory-based findings. For example, a real world study Gallup et al. (2012a) demonstrated how pedestrians would in fact ignore a gaze cue from an oncoming pedestrian, avoiding any potential joint attention on an attractive object placed between them. This clearly opposes the reflexive orienting model of an automatic system that orients gaze to any gaze cue presented. While these types of studies undoubtedly contribute significantly to our understanding of how gaze operates in a real environment, it is difficult to determine with any certainty what cognitive processes are involved in these gaze behaviours when they are examined in a real environment where controlling for all external variables is almost impossible.

Thus, it was clear that there was a need for a paradigm that investigated observers' responses to social gaze cues in a way that both better reflected real world interactions and a real world environment, whilst retaining some control of the visual information presented. The paradigm developed and tested within Chapters Two, Three and Four is a first iteration of a type of stimulus that is more realistic, presenting a gaze cue within the context of a whole body and within a real environment, but retains features of previous Posner-type tasks. For example, a single cue is presented, the person providing the cue sits in the centre of the screen and they cue to a peripheral target. However, unlike Posner-type tasks, target objects were presented within an array that occurred in a plausible way: the objects were situated on a table rather than being digitally manipulated to appear next to the person's head, or within any other spatial region of the scene. Thus, the stimulus matches with the observers' internal template of the global context of the environment (i.e. objects are affected by gravity and therefore are more likely to appear on a solid surface than in mid-air; see Malcolm & Henderson, 2009), and therefore is more likely to elicit gaze responses that would occur in a natural visual search.

Summary of Findings

Using a single gaze cue

Chapter Two presented the new paradigm with no instruction regarding person presence, but simply with an explanation of how to perform the task and a reminder that the target object should be located as quickly as possible. The results showed that person presence facilitated performance across almost all measures of search in both the search initiation and scene scanning phases. However, there was no evidence of congruency effects occurring – participants were as successful in completing the task with an incongruent gaze cue as with a congruent gaze cue. While the effects of person presence were somewhat consistent with a reflexive orienting hypothesis, the lack of congruency effects did not match previous findings in Posner-type task literature. It was postulated that the increase in scene complexity may have contributed to the differences in facilitation effects between the study in Chapter Two and previous literature, as it has been discussed by Downing et al. (2004) amongst others that the pattern of fixations across scenes is largely affected by scene context and the spatial proximity of cues and target objects.

Chapter Three investigated whether observers' eye movements would change when they were given different information regarding the presence of a person in the scenes. Previous literature (e.g. Driver et al., 1999; Ricciardelli et al., 2002) most commonly told participants that the gaze cues provided were not helpful for task completion, which is true given their spatially uninformative nature. To fully replicate previous Posner-type tasks with the newly developed paradigm, I felt it important to investigate the effects of giving participants this instruction: telling them that the gaze cues in the scenes were not helpful to the task and that the person in the scene was to be ignored. However, to test whether an instruction concerning person presence actually impacted on observers' gaze behaviour, I felt it prudent to conduct the same experiment with an instruction that suggested the person's gaze cues may be helpful in completing the task. If any differences in performance emerged between the two studies, it could be confidently assumed that the manipulation had been successful. Similar benefits to performance when a person was present in the scene were found in these studies, and in addition some effects of gaze cue congruency began to emerge. However, eye movement behaviour was consistent across both studies, which suggested that the manipulation of instruction regarding person presence had been unsuccessful. It was unclear whether this was due to participants not believing the manipulation, or whether any instruction regarding person presence made that person more salient within the scene regardless of the content of the instruction.

To provide a quantitative comparison of the effects of instruction Chapter Four presented a cross-study analysis that compared results in each of the performance measures across all three studies. This chapter found the strongest effects of instruction in the single-cue visual search task occurred within the pre-saccadic launch processing stage, evident in the first saccade latency measure. As search continued the effects of instruction diminished. It was only in the measure of error rate that any interaction between gaze cue condition and instruction condition was found, which suggests that for the majority of performance measures effects of gaze cue condition and of instruction condition occur independently. These results are surprising given the evidence that documents the impact on performance as a result of different task instructions (e.g. Fletcher-Watson et al., 2008; Itier et al., 2007; Yarbus, 1967). It is possible that the lack of instruction effect may in fact support the notion of reflexive orienting: perhaps the reflexive orienting process is so strong it cannot be overridden in search scenes such as those used in this thesis.

Using multiple gaze cues

The paradigm presented in Chapters Two through Four is a good compromise between the demands for more ecologically valid stimuli and Posner-type tasks used to date, to test how social attention is allocated in these visual search tasks. However, a key element of real world studies missing from this novel paradigm is the presence of multiple cues. While we do experience one-to-one interactions in everyday life, even reflecting on one's own experiences can highlight the frequency of occasions in which more than one gaze cue is seen at a time. Imagining walking down a busy high street or through a crowded shopping centre highlights the number of gaze cues we receive simultaneously as we navigate our environment. Real world studies can evidence this, and can report to some extent the eye movements made during this type of navigation, but to truly understand how the simultaneously-received gaze cues are processed, a simpler environment is required. The environment presented in the single-cue studies is easily adapted to accommodate two people in order to provide two gaze cues simultaneously. Thus, a first step into examining responses to multiple gaze cues can be made.

Chapter Five was the first to use simultaneously-presented multiple gaze cues in the visual search task. Having two gaze cues created an additional gaze cue condition – the *conflicting* condition where each person cued a different object. Again, this chapter tested the new stimuli with only an instruction concerning how to perform the task and no mention of the presence or absence of people in the scene. The distinct benefits of person presence seen in the previous chapters were not as clear-cut in the results of this study, and strong congruency effects were apparent, with incongruent cues causing significant disruption to participants' performance. These results provided a first indication of how multiple gaze cues are processed by observers when viewing a real world social scene. One possible explanation for these results is that these multiple cues are processed in parallel via multiple spotlights of attention, even prior to search initiation, and that this gaze cue information is accessed covertly by observers.

Chapter Six explored how the same instructions regarding the helpfulness of people presence used in Chapter Four would change observer behaviour in a multiple-cue version of the visual search task. It seems that once observers are presented with multiple social gaze cues simultaneously, the effects of instruction become much stronger. When told to ignore the people in the scene, congruency effects were strongest in the search initiation phase where the congruent gaze cue condition elicited better participant performance than the incongruent condition. However, when told the people in the scene may be helpful in finding the target the effects of congruency occurred in a different stage of search. No congruency effects emerged during search initiation, but were clearly apparent in the scene scanning phase where congruent gaze cues most strongly facilitated participants' performance. These findings fit much more closely with the effects of task instruction predicted by Fletcher-Watson et al. (2008), Itier et al. (2007) and Yarbus (1967), amongst others.

Chapter Seven documented the quantitative analysis of the effects of task instruction in an omnibus chapter that compared all three multiple-cue studies, examining each performance measure separately. When compared, the effects of instruction were surprising. Analyses showed that an instruction suggesting people presence may be helpful in finding the target object actually resulted in significant deterioration in performance across several measures. One possible explanation for these results is that the instructions, particularly in the helpful instruction condition, generate an attentional set that causes them to prioritise objects cued by the people in the scene. Since these cues are non-predictive this type of attentional set will only benefit participants on half of trials. In the remaining trials they have to process these gazed-at objects before continuing search through the rest of the array, thus resulting in slower and less efficient searches. This may also account for the interactions found between gaze cue condition and instruction condition in the response time and error rate measures. As discussed by Summerfield and Egner (2009), expectations guide visual processing, and in the current research an instruction that makes people presence salient generates an expectation which may inform the attentional set. Alternatively, it may be possible that an implicit Theory of Mind system causes increased cognitive load in any people-salient instruction condition, which might explain the general deterioration in performance seen when either a helpful or unhelpful instruction is given. Whatever the cause, it is clear from the findings in this chapter that task instruction has a clear, demonstrable effect when observers receive more than one gaze cue simultaneously.

What was unclear from the studies that presented multiple cues that were spatially uninformative was how observers processed conflicting gaze cue information. Given that in the real world we are presented with a variety of gaze cues, some more relevant to our task than others, I felt a good way to gain some insight into the cognitive processes behind this cue-selection would be to present observers with spatially informative cues, comprised of the conflicting gaze cue condition scenes. This type of approach was also highlighted by Langton and Bruce (1999) as a means by which the social signals used by observers could be better understood. Chapter Eight explored the effects of manipulating gaze cue sender reliability across all three instruction conditions. Observers either received cues that reliably predicted target location on 70% of trials, or cued the distractor on 70% of trials. The results showed that the benefits of people presence to observers' performance in the task persist when the cues they are given are spatially informative. There was strong evidence that overt gaze-seeking increased when observers were given an instruction suggesting the people may be helpful in finding the target object, which suggests the instruction manipulation was successful, and that observers believed what they were told about the presence of people in the scene. Reliability effects were not as strong as the congruency effects seen in previous chapters, and when reliability effects are present it seems that the unreliable condition, where the unreliable gaze cue sender cues the target, results in the best performance. It is possible that this counter-intuitive result is the effect of participants disregarding any social gaze cue information when they find it unreliable.

A summary of research findings

Overall, the studies documented within this thesis explore the different ways in which observers utilise gaze cues presented in visual search tasks, and how their eye movements vary depending on task instruction, gaze cue sender reliability, and the number of people present in a scene. This research demonstrates that some of the effects found Posner-type-task literature persist when a similar task is used in a more realistic environment, but that gaze behaviour seems to change considerably with the introduction of more than one gaze cue. The research contained in this thesis then stands as a good example for why it is important to think critically about the environments in which we encounter social gaze cues, and to ensure the materials and stimuli used to investigate these phenomena reflect the types of social gaze experienced in the real world.

Implications of This Research

Contributions to the debate of reflexive orienting

This research adds to our understanding of reflexive orienting in response to social gaze cues, as well as suggesting the sort of cognitive model that might be employed to process multiple gaze cues. There is a wealth of established literature that documents the seemingly inherent and automatic behavioural response to a gaze cue: even when we know this cue is unhelpful, our attention (indicated by an eye movement) is drawn to the area cued by another person's gaze (e.g. Driver et al., 1999; Friesen & Kingstone, 1998, 2003a, 2003b; Kuhn & Benson, 2007; Ricciardelli et al., 2002). It has been suggested that this response is due to the special status of eyes: when presented alongside other stimuli eyes are preferentially-fixated over everything else (e.g. Birmingham et al., 2007; Frith, 2008). When eye information is unavailable, gaze information is inferred from head direction and objects cued by this holistic understanding of gaze – that eyes and head normally point in the same direction – elicit the same reflexive orienting from observers (e.g. Zwickel & Vō, 2009). However, studies examining the automatic orienting of attention in response to a centrally-presented gaze cue have received criticism when applying their findings to real world behaviour because the stimuli they use is so far removed from real world experience (Risko et al., 2012).

The studies contained in Chapters Two and Three (and further analysed in Chapter Four) present a more realistic environment for studying conceptually similar stimuli. As has already been stated, scenes show a centrally-presented figure who cues to their left or right, which is a similar format to Posner-type studies. To some extent, the findings of these chapters support a reflexive orienting hypothesis. In essence, reflexive orienting is evidence of gaze seeking and following – the observer follows the gaze of the face and so is faster to find the target when the cue is congruent, but slower when the cue is incongruent. When given a single gaze cue, person presence – regardless of cue type – facilitated performance. This suggests that to some extent the cue being given by the person in the scene is being processed by the observer, though the lack of overt gaze-seeking may indicate this information is being accessed covertly; processing gaze covertly has been evidenced in previous research by Knoeferle and Kreysa (2012), and Macdonald and Tatler (2013a).

In many of the performance measures in the single cue studies, across all instruction conditions, there were numerical differences in performance between the incongruent and congruent gaze cue conditions, but the difference was not sufficient to be statistically significant. In the multiple cue studies presented in Chapters Five, Six and Seven, participants still garnered the benefits of people presence in the scene, again suggesting that the gaze information provided is being processed and applied to the task. Furthermore, once two people were present in the scene, an incongruent gaze cue strongly disrupted performance, which mirrors what occurred when an incongruent cue is presented in a Posner-type task. Therefore, this evidence reinforces the idea of reflexive orienting to cues provided.

However, while some congruency effects were apparent, they do not match the consistent effects documented in Posner-type tasks. The current research utilises a wider range of performance measures than earlier research, but even if the response times measure is taken in isolation (to match the measure most commonly used in previous research) the congruency effects do not match. Effects presented by Friesen and Kingstone (2003) and Driver et al. (1999), as two examples amongst many, demonstrate that a cognitive cue results in a statistically significant faster response time than the response time that occurs when the participant is given an incongruent cue. This is not found in the current research. Response times may be faster for gaze cues in people present scenes as opposed to people absent scenes, but there was not the same degree of difference between people present conditions.

One explanation that may account for why these studies cannot so clearly evidence the benefits of congruent gaze cues seen in previous literature is that the realistic environment means that the eyes take up a much smaller proportion of the visual field compared to stimuli containing only a centrally-presented face (Fletcher-Watson et al., 2008). When given the latter type of cue, the eyes are much clearer and thus the direction in which they are gazing is more readily apparent. Although research by Birmingham et al. (2009) and Perrett et al. (1992) has demonstrated that humans process social information in a hierarchical format, and that when eye gaze information is absent other social cues (e.g. head direction) will be the next most preferred stimulus, it is unlikely that head direction can convey the same degree of facilitatory benefits. If eyes are truly a special stimulus, (as suggested by Birmingham et al., 2007, 2008, 2009; Downing et al., 2004; Perret et al., 1992; amongst others) it may be possible that when this information is not clearly available, and participants must use other information, the benefits derived are not as great. This would account to some degree for evidence gathered from the studies within this thesis somewhat supporting a reflexive orienting hypothesis, but not completely.

The mixed effects of task instruction also present some difficulty for interpreting the results found in these chapters by means of a reflexive orienting hypothesis alone. It is clear that the single cue studies emulate the same lack of instruction effect as the original Posner-type studies; that is, instructing participants to ignore the gaze cue presented had no effect on their eye movement behaviour in response to that cue. In this way, the single cue studies support findings of previous Posner-type-task studies. It was suggested by Langton et al. (2000) that cells in the inferior temporal (IT) cortex of the right hemisphere were dedicated to processing gaze direction, thus were triggered when presented with gaze stimuli and act as the origin for reflexive orienting. This could explain why in instances where participants' are told that gaze is counter-predictive (as in Driver et al., 1999), which means the target would appear on the opposite side to which the face gazed, participants still oriented their gaze first to the cued side of the screen. This indicates that the orienting triggered by cells in the IT cortex operate independently of top-down processes. However, Ristic and Kingstone (2005) were able to demonstrate that the reflexive orienting must originate in part from top-down controls, and that these controls take a role in activating the IT cortex's involvement in gaze processing. When the authors presented participants with the same stimulus and told them in one condition that the stimulus was a car, and in another that it was a pair of eyes, only the eyes triggered any reflexive orienting. However, when participants were shown both conditions, the ordering of the instruction changed their viewing behaviour. Being told the stimulus was a car first meant that reflexive orienting occurred only for the eyes condition, but if the participants were told the stimulus was a pair of eyes first, and then a car, the reflexive orienting could not be inhibited. Ristic and Kingstone (2005) state that these findings suggest we possess top-down mechanisms that activate the IT cortex cells when a stimulus is perceived as a face, and that once the IT cortex is activated it cannot be overridden; once a stimulus is seen to have eyes the way it is perceived cannot be changed. In terms of the research documented within this thesis, this evidence would suggest that in single cue studies, IT cortex cells are activated once the stimuli presented are perceived to contain a face, and thus the reflexive orienting processes are engaged.

Yet this explanation does not translate well to real world research, nor the multiple cue studies discussed from Chapter Five onwards. When two people are present in the scene there are clear effects of instruction: levels of overt gaze-seeking changed depending on what the participant was told about the nature of the gaze cues, and their performance across a range of measures was affected by the instruction condition they were in. This is much more in-keeping with the research documented by Yarbus (1967), Itier et al. (2007), and Gallup et al. (2012a), who demonstrated that the instruction given to a participant changes the way in which they view the scene. This is certainly the case in the multiple cue studies reported here. While the simplified, but realistic, visual search task does not emulate the sort of real world interaction with other people discussed in Gallup et al.'s (2012a) work for one, it does create a more dynamic gaze cuing scenario than single cue paradigms, and the evidence produced from these studies support findings that show task instruction affects eye movement behaviour.

It seems then that when the task presents the same stimuli, but for the addition of a second person, results take quite a step away from studies that support the reflexive orienting model. Since the only difference is moving from a single cue to a multiple cue, it must be concluded that when presented with multiple cues the way in which gaze information is responded to changes. Whether there are one or many factors involved in this processing shift cannot be answered definitively from the work discussed here. However, what this work does demonstrate is that observers' responses to multiple gaze cues are not best explained by a reflexive orienting model. This does not mean this model is incorrect: as has been discussed above, the single cue studies using a realistic paradigm do seem to evidence some reflexive orienting behaviour. While the results do not conform completely to previous evidence, this may be due to the increased complexity of the scene and/or task, the shorter number of trials given to participants, or the longer duration for which the scene was available. All these findings suggest is that there are other, more complex processes at work when interpreting and utilising multiple gaze cues.

Alternatively, it is possible that the improvement in performance in person-present scenes compared to person-absent scenes is the result of social facilitation. First documented by Triplett (1898), social facilitation is the phenomenon in which a person's performance in a task is improved simply by the mere presence of another person. Whereas some studies rely on the actual presence of another person during the task (e.g. Allport, 1920; Böckler, Knoblich & Sebanz, 2012; Sebanz, Bekkering & Knoblich, 2006), others have demonstrated that social facilitation occurs even when participants believe another unseen person is viewing the same stimuli, but have no proof of this (e.g. Richardson, Hoover & Ghane, 2008; Richardson et al., 2012). There is a strong body of evidence that suggests by having a person present in the scene, participants' performance should improve – regardless of the gaze cue provided. While the reflexive orienting model struggles to account for the lack of congruency effects observed in the single cue studies, evidence from Markus (1977) and others discussed in Zajonc (1965) suggest that when engaging in a complex or unfamiliar task the presence of person can detrimentally affect participant performance. In Markus' (1977) study, when participants dressed or undressed in unfamiliar clothing with either a passive or active observer, they were much slower than when they dressed or undressed alone. However, when dressing and undressing in familiar clothing, an observer facilitated their performance. Considering this in terms of the evidence documented within this thesis, it is possible that participants benefit from a having a person in the scene when searching for a familiar object, but are hindered by person presence when searching for an object that is unfamiliar. Thus, it is likely that it is a combination of reflexive orienting and social facilitation models that best account for the results observed in the single cue studies.

Implications for multiple gaze cue processing

If a more complex strategy for processing multiple cues is being suggested, what form might this take? As has been discussed in previous chapters, it is possible that having two gaze cues rather than one necessitates the utilisation of multiple spotlights of attention deployed and processed in parallel (see Pylyshyn & Storm, 1998; Scholl et al., 2001). Evidence from fMRI studies conducted by McMains and Somers (2004, 2005) clearly demonstrate neural activity during a visual monitoring task that shows two spatially distinct areas within the visual cortex processing the visual information simultaneously. It would be sensible then to imagine that a cognitive process employed for interpreting multiple cues simultaneously would utilise some form of multiple spotlight model of attention, because even in conditions in the current research where both people cue the same object, these cues are from spatially distinct areas and thus require discrete processing. Applying this to navigation of a real world environment where multiple cues must be assessed for their usefulness in terms of the current task (e.g. walking through a crowd), a multiple spotlight model of attention is logical; processing all these different cues serially would be highly inefficient, if not impossible to do. However, it is equally possible that we may disregard much of the stimuli within the environment, focusing only on what is task relevant in each situation.

What does seem clear is that people presence was highly beneficial to performance in the search task, regardless of gaze cue condition. Perrett et al. (1992) describe that amongst all social stimuli, gaze is at the top of the stimulus hierarchy, and multiple studies have documented the preference for looking at other peoples' eyes over other stimuli (e.g. Birmingham et al., 2007; Birmingham et al., 2008a; Castelhano et al., 2007; Emery, 2000; Langton et al., 2000). Given this well-documented preference for gaze, it is unsurprising that gaze cues impact on search, even when these cues are incongruent with target location. While there is variation in response time as a result of task instruction, people present scenes result in faster searches than people absent scenes. As has been discussed in previous chapters, it is likely that the gaze cues provided within the search scenes contribute to the participants' internal target template (see Malcolm & Henderson, 2009). This template contains an individual's internal representation of an object's features, and according to Wolfe et al. (2004) these target features weight spatial locations in the scene according to the degree to which they match with the template, guiding subsequent eye movements. One possible explanation for the facilitatory effects of people presence in the search task is that gaze cues provided in people present scenes are additional cues that weight attention to multiple locations within the scene (Awh & Pashler, 2000; Casteillo & Umiltà, 1990). This would facilitate search even if participants had difficulty determining the specific object being cued by the gaze cue sender – more weighting to a particular spatial location guides eye movements (and therefore search) to this area, thus ruling out objects out with this area as potential targets. Considering the evidence described above, which suggests the high importance of gaze information in cognitive processing of scenes, it is logical to assume that gaze would contribute further weighting to a spatial location within the scene. The benefit of people presence to search makes sense: if participants follow the gaze cues the search array of objects is reduced from 15 to approximately two (allowing for occasions where participants may not be able to define the specific target cued by gaze). Even if participants follow an incongruent gaze cue first, their search would still be faster than people absent scenes where all 15 objects are equally likely to be the target – indeed this is evidenced in the results of Chapters Five through Seven.

Considering the evidence gathered through the current research, the strategy for processing multiple gaze cues simultaneously also takes note of task instruction. Data from the overt gaze-seeking measure in Chapters Five to Seven clearly show that participants apply what they have been told regarding people presence to their visual search strategy. If they did not, there would be no significant increase in overt gaze-seeking in the helpful instruction condition compared to both the no instruction and unhelpful instruction conditions. As was evidenced by Yarbus (1967), different instructions result in different eye movements. How does this evidence translate to a model of multiple gaze cue processing? Task instruction may be used to discriminate helpful stimuli. For example, when participants are told that the people in the scene are not helpful to the task, response times (to take one performance measure as an example) in this condition were consistent with the no instruction condition, and both were faster than the helpful instruction condition. Considering the top-down processes discussed above, which Ristic and Kingstone (2005) said controlled the activation of IT cortex cells responsible for processing gaze, perhaps a helpful instruction allows these processes to exert control over IT cortex cell activation and inhibit reflexive orienting or other gaze processing responses. This would allow gaze information to be disregarded, and in doing so, reduce the amount of information that requires processing in order to find the target. That individuals can regulate their attention to be sensitive to different stimuli depending on task demands has already been discussed in previous research (Folk et al., 1992). By the same reasoning, this would also account for why performance in the helpful instruction condition deteriorates: participants have selected the gaze cues as relevant and important to task completion and thus dedicate cognitive resources to processing them. However, since cues are spatially uninformative they are not actually deriving any benefit from this additional processing.

It is also possible that when processing multiple gaze cues observers employ strategic gaze-following behaviour. Ricciardelli et al. (2013) revisited previous research (Ricciardelli et al., 2002) in which a digitised face was used to provide centrally-presented gaze cues to peripheral targets, but the authors added distracting gaze cues to empty spatial locations as well as varying instructions given to participants. Participants were instructed to look at one of two targets while the face cued a task-relevant target, a task-irrelevant target, or an empty location. Ricciardelli et al. (2013) found that participants never followed gaze cues to the empty location, but they did make more errors when presented with a distracting incongruent gaze cue. The authors suggest these findings document a strategic gaze-following strategy, which has elements of both automatic and goal-driven gaze orienting. Thus, processing multiple cues in the real world is to some degree reflexive, but we can choose to ignore cues if they are irrelevant to our current task. This is supported by evidence from Fletcher-Watson et al. (2008) who showed that the proportion of fixations on a person’s face dropped significantly when participants were asked to discriminate the gender of the person in the scene rather than to freely view the scene. Gaze is irrelevant to gender discrimination, and participants adapted their viewing behaviour appropriately. Similarly in the current research, an unhelpful instruction resulted in significantly fewer fixations on either person's face than when they were given a helpful instruction. Cumulatively this evidence suggests that the processing of multiple gaze cues is more selective than when viewing a single cue, with greater influence of goal-driven mechanisms that reduce reflexive gaze-following and instead selectively attending only the gaze cues that are useful to the task.

How does the reliability of multiple cues encountered influence gaze cue processing? Langton and Bruce (1999) were among the first to vary the reliability of gaze cues presented to participants; that is, they varied the rate at which the gaze cue would predict target location. They employed three different cue validity rates: 25%, 50% and 75%. Their results showed that uninformative or to-be-ignored cues still resulted in improved performance, but only when cues appeared 100 ms before target onset. Increasing the reliability of the gaze cue to predict target location on 75% of trials did not result in any greater facilitation of performance. Similar results were found in the studies documented in Chapter Eight. While there are still benefits of people presence to performance, there are rarely any differences in performance when the target is cued by the reliable gaze cue sender or by the unreliable gaze cue sender. This means that the evidence currently available cannot inform how reliability impacts on the processing of multiple gaze cues, other than to indicate any gaze cue (i.e. people presence) is better than none.

Perhaps the difficulty in eliciting reliability effects is due to the way in which reliability is attributed in the real world, compared to how it is attributed in laboratory-based paradigms. In the lab, reliability is attributed to the degree to which a face or a person cues the target object. However, in real world interaction, reliability has different connotations. We associate reliability with personal characteristics, such as trustworthiness. Research by King et al. (2011) showed how manipulating the perceived trustworthiness of a gaze cue sender by providing participants with vignettes about the people they would see in the scenes presented to them, affected the degree to which a participant preferred an object cued by the person, and the speed with which they categorised that object.

While there is much more research required to develop an accurate model of how we process multiple gaze cues, the research contained within this thesis makes a comprehensive start. It has developed and tested a paradigm within which observers are presented with multiple cues simultaneously in a realistic environment, making the stimuli more ecologically valid whilst still retaining control of what participants see and when. The evidence presented suggests that gaze cue congruency has a stronger effect in multiple gaze cue scenes than in single cue scenes, with two incongruent gaze cues being much more distracting than just one incongruent cue, but without always finding the expected benefits of congruent cues. The results also demonstrate that task instruction impacts on participants' performance, which is much more in-keeping with findings from real world research. However, it is still unclear how observers process gaze cue sender reliability, or if any differences in reliability are even noticed by observers.

Future Directions

The difficulty in determining how observers process gaze cue sender reliability highlights one of several potential directions for future research based on the work presented in this thesis. I have shown that reflexive orienting seems to persist in single cue studies that use realistic stimuli, so future research should focus on variations of the multiple cue paradigm. One of the most important questions that remains unanswered is how observers process cues that offer different levels of reliability in terms of cuing the target object. It is important to consider how people determine reliability in the real world when studying this in the laboratory. If we deem someone as reliable because they are perceived to be trustworthy (as in King et al., 2011), then the studies described in Chapter Eight could be repeated, but with an additional manipulation that confers different degrees of trustworthiness to the people in the scene. Combining this with an actual manipulation of reliability may begin to elicit stronger reliability effects.

Additionally, it would be useful to examine whether the effects documented in response to spatially-uninformative multiple gaze cues persist when observers are shown dynamic gaze cues. Given the differences in responses found in laboratory-based studies versus those conducted in the real world, it is possible that static cues do not elicit the same types of responses as dynamic cues. Short video clips of natural gaze cues, where the people in the scene begin by looking straight ahead and then to the object they are cuing would be far more representative of real world social attention, and thus have much greater ecological validity. If the same eye movement behaviour occurred, it would be possible to conclude that the novel paradigm presented within this thesis achieves an accurate representation of real world gaze. However, should the facilitation and congruency effects disappear or change, this would suggest that only dynamic cues can accurately represent real world social attention. This would be particularly important for further research into social attention, and would set a gold standard for ecologically valid stimuli.

Since these studies are among the first to document how multiple cues impact observer eye movement behaviour in a visual search task, they provide a baseline to which other populations can be compared. Taking a paradigm first tested in a normal population and applying it to a special population, for example those with autism, allows us to better understand the differences in how one population views the world compared to another. For example, Freeth, Foulsham and Kingstone (2013) examined participants' viewing behaviours during one-to-one social interactions conducted either via video or in person. The participants had to answer some questions while the level of eye contact given by the experimenter was manipulated to be either direct or averted. All participants self-reported autistic traits using the Autism-spectrum Quotient (AQ) questionnaire. Results showed that a higher score on the AQ – indicating more autistic traits – correlated with a smaller proportion of viewing time of the person shown in the video condition, which replicated previous findings showing individuals with autism are less likely to look at people when watching videos (e.g. Klin et al., 2002; von Hofsten, Uhlig, Adell & Kochukhova, 2009). Similar to Freeth et al.'s (2013) work, it is possible that the current studies could be adapted to explore the differences in how special populations interpret multiple cues. This type of comparison allows better understanding of the difficulties others may have in interpreting social cues, and thus indicate ways in which they can be better supported. This would be a particularly valuable future direction for the research begun here.

Conclusions

This thesis aimed to develop and test a paradigm that built on the foundation of research using Posner-type tasks whilst retaining more ecological validity by being more representative of a real world experience of gaze. The first iteration of this paradigm used a single cue, and it is this version of the paradigm that most closely resembles the previous Posner-type task literature. Participants completed the visual search task under different instruction conditions that manipulated the perceived helpfulness of the gaze cues provided in people present scenes. Results across these studies showed that while the presence of a person in the scene facilitated performance across several different measures, there were no significant effects of gaze cue congruency. For the most part these results supported a reflexive orienting model of gaze, and it was suggested that the difference in findings between the current studies and previous research could be due to a number of factors regarding increased stimulus complexity. It was discussed that in addition to the reflexive orienting model, additional consideration of social facilitation may best account for the results observed in these studies. When a second cue was introduced, thus presenting participants with multiple gaze cues simultaneously, observers' responses changed quite considerably. Clear congruency effects began to emerge, and the instructions given to participants regarding people presence had clear impact on performance across the different gaze cue conditions. These findings cannot be explained by a reflexive orienting model of gaze. A model of multiple gaze cue processing was discussed, and it was hypothesised that this may include a multiple spotlight model of attention and would take note of task instruction. However, the findings regarding the impact of manipulating the reliability of the gaze cue sender on observer eye movement behaviour were inconclusive. It was suggested that how reliability is inferred by participants may be different from how it is attributed in the laboratory, and that future research should aim to consider how reliability is attributed in the real world and make attempts to match this process in laboratory-based testing. Furthermore, it was stated that this paradigm should be adapted to include dynamic gaze cues. This would allow the determination of how accurately the paradigm represents real world social gaze, and to what degree cues need to be dynamic to truly emulate real world gaze. Finally, the potential for using the findings documented within this thesis as a baseline for comparison to responses in social gaze was discussed, and it was suggested that this comparison could provide useful insights into social difficulties of those in special populations. Cumulatively, this research not only contributes to the discussion of a reflexive orienting model of gaze, but adds a new dimension to this knowledge by studying multiple social gaze cues from a number of different perspectives using different task instructions and different levels of gaze cue sender reliability, and highlights several important directions for future research in this area.

References

Abrams, R. A., & Christ, S. E. (2003). Motion Onset Captures Attention. *Psychological Science, 14,* 427-432.

Aimola Davies, A. M., Waterman, S., White, R. C., & Davies, M. (2013). When you fail to see what you were told to look for: Inattentional blindness and task instructions. *Consciousness and Cognition, 22,* 221-230.

Allport, F. H. (1920). The influence of the group upon association and thought. *Journal of Experimental Psychology, 3,* 159-182.

Apperly, I. A., & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review, 116,* 953-970.

Argyle, M., & Dean, J. (1965). Eye-Contact, Distance and Affiliation. *Sociometry, 28,* 289-304.

Awh, E., & Pashler, H. (2000). Evidence for split attentional foci. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 834-846.

Baron-Cohen, S., Baldwin, D. A., & Crowson, M. (1997). Do children with autism use the speaker's direction of gaze strategy to crack the code of language? *Child Development, 68,* 48-57.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition, 21,* 37-46.

Baron-Cohen, S., Wheelwright, S., & Joliffe, T. (1997). Is There a “Language of the Eyes”? Evidence from Normal Adults, and Adults with Autism or Asperger Syndrome. *Visual Cognition, 4,* 311-331.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *Lme4: linear mixed-effects models using eigen and s4.* R package version 1.1-6. Retrieved from: http://CRAN.R-project.org/package=lme4

Bayliss, A. P., & Tipper, S. P. (2006). Predictive Gaze Cues and Personality Judgements: Should Eye Trust You? *Psychological Science, 17,* 514-520.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2007). Why do we look at people’s eyes? *Journal of Eye Movement Research, 1,* 1-6.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008a). Social attention and real-world scenes: The roles of action, competition and social content. *The Quarterly Journal of Experimental Psychology, 61,* 986-998.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008b). Gaze selection in complex social scenes. *Visual Cognition, 16,* 341-355.

Birmingham, E., Bischof, W. F., & Kingstone, A. (2009). Get real! Resolving the debate about equivalent social stimuli. *Visual Cognition, 17,* 904-924.

Birmingham, E., & Kingstone, A. (2009). Human Social Attention: A New Look at Past, Present, and Future Investigation. *Annals of the New York Academy of Sciences, 1156,* 118-140.

Böckler, A., Knoblich, G., & Sebanz, N. (2012). Effects of a coactor's focus of attention on task performance. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 1404-1415.

Brown, V., Huey, D., & Findlay, J. M. (1997). Face detection in peripheral vision: do faces pop out? *Perception*, *26*, 1555-1570.

Buschman, T. J., & Miller, E. K. (2010). Shifting the spotlight of attention: evidence for discrete computations in cognition. *Frontiers in Human Neuroscience, 4,* 1-9.

Buswell, G. T. (1935). *How people look at pictures.* Chicago, IL: University of Chicago Press.

Cañadas, E., & Lupiáñez, J. (2012). Spatial interference between gaze direction and gaze location: A study on the eye contact effect. *The Quarterly Journal of Experimental Psychology*, *65*, 1586-1598.

Cannon, L. W., Higginbotham, E., & Leung, M. L. A. (1988). Race and Class Bias in Qualitative Research on Women. *Gender and Society, 2,* 449-462.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63,* i-vi, 1-143.

Carrasco, M., & McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 5363-5367.

Castelhano, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception, & Psychophysics, 72,* 1283-1297.

Castelhano, M. S., & Henderson, J. M. (2007). Initial Scene Representations Facilitate Eye Movement Guidance in Visual Search. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 753-763.

Castelhano, M. S., Wieth, M., & Henderson, J. M. (2007). I See What You See: Eye Movements in Real-World Scenes Are Affected by Perceived Direction of Gaze. In L. Paletta & E. Rome (Eds.). *Attention in Cognitive Systems. Theories and Systems From an Interdisciplinary Viewpoint.* (pp. 251-262). Berlin, Germany: Springer-Verlag.

Castiello, U., & Umiltà, C. (1992). Splitting Focal Attention. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 837-848.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception-behaviour link and social interaction. *Journal of Personality and Social Psychology, 76,* 893-910.

Crosby, J. R., Monin, B., & Richardson, D. (2008). Where Do We Look During Potentially Offensive Behaviour? *Psychological Science, 19,* 226-228.

Downing, P. E., Dodds, C. M., & Bray, D. (2004). Why does the gaze of others direct visual attention? *Visual Cognition, 11,* 71-79.

Driver, J., Davis, G., Ricciardelli, P., Kidd, P., Maxwell, E., & Baren-Cohen, S. (1999). Gaze Perception Triggers Reflexive Visuospatial Orienting. *Visual Cognition, 6,* 509-540.

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision, 8*, 1-19.

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioural Reviews, 24*, 581-604.

Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception & Psychophysics, 40,* 225-240.

Eriksen, C. W., & Yeh, Y. Y. (1985). Allocation of attention in the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 11*, 583-597.

Ewbank, M. P., Jennings, C., & Calder, A. J. (2009). Why are you angry with me? Facial expressions of threat influence perception of gaze direction. *Journal of Vision, 9*, 1-7.

Feinman, S. (1982). Social Referencing in Infancy. *Merrill-Palmer Quarterly, 28,* 445-470.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception, 37,* 571-583.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance*, *18,* 1030-1044.

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstone, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition, 117,* 319-331.

Freeth, M., Foulsham, T., & Kingstone, A. (2013). What Affects Social Attention? Social Presence, Eye Contact and Autistic Traits. *PLoS ONE, 8,* e53286.

Friesen, C. K., & Kingstone, A. (1998). The eyes have it! Reflexive orienting is triggered by nonpredictive gaze. *Psychonomic Bulletin & Review, 5,* 490-495.

Friesen, C. K., & Kingstone, A. (2003a). Abrupt onsets and gaze direction cues trigger independent reflexive attentional effects. *Cognition, 87,* B1-B10.

Friesen, C. K., & Kingstone, A. (2003b). Covert and overt orienting to gaze direction cues and the effects of fixation offset. *NeuroReport, 14,* 489-493.

Friesen, C. K., Ristic, J., & Kingstone, A. (2004). Attentional Effects of Counterpredictive Gaze and Arrow Cues. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 319-329.

Frith, C. D. (2008). Social cognition. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences, 363,* 2033-2039.

Frith, C. D., & Frith, U. (2007). Social Cognition in Humans. *Current Biology, 17,* 724-732.

Funes, M. J., Lupiáñez, J., & Milliken, B. (2007). Separate mechanisms recruited by exogenous and endogenous spatial cues: Evidence from a spatial Stroop paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 33,* 348-362.

Gallup, A. C., Chong, A., & Couzin, I. D. (2012a). The directional flow of visual information transfer between pedestrians. *Biology Letters, 8,* 520-522.

Gallup, A. C., Hale, J. J., Sumpter, D. J. T., Garnier, S., Kacelnik, A., Krebs, J. R., & Couzin, I. D. (2012b). Visual attention and the acquisition of information in human crowds. *Proceedings of the National Academy of Sciences of the United States of America, 109,* 7245-7250.

Geng, J. J., & Behrmann, M. (2005). Spatial probability as an attentional cue in visual search. *Perception and Psychophysics, 67,* 1252-1268.

Gibson, E. J., & Walk, R. D. (1960). The "Visual Cliff". *Scientific American, 202,* 67-71.

Godijn, R., & Theeuwes, J. (2002). Programming of endogenous and exogenous saccades: evidence for a competitive integration model. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 1039-1054.

Green, J. J., Gamble, M. L., & Wolfdorff, M. G. (2013). Resolving conflicting views: Gaze and arrow cues do not trigger rapid reflexive shifts of attention. *Visual Cognition, 21,* 61-71.

Greene, D. J., Mooshagian, E., Kaplan, J. T., Zaidel, E., & Iacoboni, M. (2009). The neural correlates of social attention: automatic orienting to social and non-social cues. *Psychological Research, 73,* 499-511.

Guerin, B. (2010). Social Facilitation. In I. B. Weiner, & W. E. Craighead (Eds.). *Corsini Encyclopedia of Psychology, 4th Edition* (pp. 1-2). New York: Wiley.

Gullberg, M., & Holmqvist, K. (2002). Visual Attention towards Gestures in Face-to-Face Interaction vs. on Screen. *Gesture and Sign Language in Human-Computer Interaction: Lecture Notes in Computer Science, 2298*, 206-214.

Haider, H., & Frensch, P. A. (1999). Information reduction during skill acquisition: The influence of task instruction. *Journal of Experimental Psychology: Applied, 5,* 129-151.

Haji-Abolhassani, A., & Clark, J. J. (2014). An inverse Yabus process: Predictive observers' task from eye movement patterns. *Vision Research, 103,* 127-142.

Hayhoe, M. M., & Ballard, D. (2005). Eye movements in natural behaviour. *Trends in Cognitive Sciences, 9,* 188-194.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision, 3,* 1-24.

Healey, C. G., & Enns, J. T. (2011). Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Transactions on Visualization and Computer Graphics, 42,* 1-20.

Henderson, J. M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50,* 243-271.

Henderson, J. M., McClure, K. M., Pierce, S., & Schrock, G. (1997). Object identification without foveal vision: Evidence from an artificial scotoma paradigm. *Perception and Psychophysics, 59,* 323-346.

Henderson, J. M., Williams, C. C., & Falk, R. J. (2005). Eye movements are functional during face learning. *Memory and Cognition, 33,* 98-106.

Henrich, J., & Gil-White, F. (2001). The evolution of prestige: Freely conferred status as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behaviour, 22,* 1-32.

Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans Have Evolved Specialized Skills of Social Cognition: The Cultural Intelligence Hypothesis. *Science, 317,* 1360-1366.

Herschler, O., & Hochstein, S. (2005). At first sight: a high-level pop-out effect for faces. *Vision Research, 46,* 3028-3035.

Hietanen, J. K. (2002). Social attention orienting integrates visual information from head and body orientation. *Psychological Research, 66,* 174-179.

von Hofsten, C., Uhlig, H., Adell, M., & Kochukhova, O. (2009). How children with autism look at events. *Research in Autism Spectrum Disorders, 3,* 556-569.

Hori, E., Tazumi, T., Umeno, K., Kamachi, M., Kobayashi, T., Ono, T. & Nishijo, H. (2005). Effects of facial expression on shared attention mechanisms. *Physiology and Behaviour, 84,* 397-405.

Horowitz, T. S., & Wolfe, J. M. (2001). Search for multiple targets: Remember the targets, forget the search. *Perception and Psychophysics, 63,* 272-285.

Huckfeldt, R. (2001). The Social Communication of Political Expertise. *American Journal of Political Science, 45,* 425-438.

Hunt, A. R., & Kingstone, A. (2003). Covert and overt voluntary attention: linked or independent? *Cognitive Brain Research, 18,* 102-105.

Itier, R. J., Villate, C., & Ryan, J. D. (2007). Eyes always attract attention but gaze orienting is task-dependent: Evidence from eye movement monitoring. *Neuropsychologia, 45,* 1019-1028.

James, W. (1890/1950). *The Principles of Psychology.* New York: Dover.

Jonides, J. (1981). Voluntary vs. automatic control over the mind's eye's movement. In J. B. Long & A. D. Baddeley (Eds.). *Attention and Performance IX* (pp. 187-203). Hillsdale, NJ: Elbaum

King, D., Rowe, A., & Leonards, U. (2011). I Trust You; Hence I Like the Things You Look At: Gaze cueing and sender trustworthiness influence object evaluation. *Social Cognition, 29,* 476-485.

Kingstone, A., Smilek, D., Ristic, J., Friesen, C. K., & Eastwood, J. D. (2003). Attention, Researchers! It is Time to Take a Look at the Real World. *Current Directions in Psychological Science, 12,* 176-180.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46,* 1762-1776.

Kleigl, R., Hohenstein, S., Yan, M., & McDonald, S. A. (2012). How preview space/time translates into preview cost/benefit for fixation durations during reading. *The Quarterly Journal of Experimental Psychology, First,* 1-20.

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry, 59,* 809-816.

Knapp III, W. H., & Abrams, R. A. (2012). Fundamental differences in visual search with verbal and pictorial cues. *Vision Research, 71,* 28-36.

Knoeferle, P., & Kreysa, H. (2012). Can speaker gaze modulate syntactic structuring and thematic role assignment during spoken sentence comprehension? *Frontiers in Psychology, 3,* 538.

Kobayashi, H., & Koshima, S. (1997). Unique morphology of the human eye. *Nature, 387,* 767-768.

Kuhn, G. & Benson, V. (2007). The influence of eye-gaze and arrow pointing distractor cues on voluntary eye movements. *Perception and Psychophysics, 69,* 966-971.

Laidlaw, K. E. W., Foulsham, T., Kuhn, G., & Kingstone, A. (2011). Potential social interactions are important to social attention. *Proceedings of the National Academy of Sciences of the United States of America, 108,* 5548-5553.

Land, M. F. (1999). Motion and vision: why animals move their eyes. *Journal of Comparative Physiology A, 185,* 341-352.

Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41,* 3559-3565.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living: making a cup of tea. *Perception, 28,* 1311-1328.

Land, M. F., & Tatler, B. W. (2009). *Looking and Acting: Vision and Eye Movements in Natural Behaviour.* Oxford, UK: Oxford University Press.

Langton, S. R. H. (2000). The mutual influence of gaze and head orientation in the analysis of social attention direction. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 53,* 825-845.

Langton, S. R. H., & Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition, 6,* 541-568.

Langton, S. R. H., Watt, R. J., & Bruce, V. (2000). Do the eyes have it? Cues to the direction of social attention. *Trends in Cognitive Sciences, 4,* 50-58.

Leder, H., Tinio, P. P. L., Fuchs, I. M., & Bohrn, I. (2010). When attractiveness demands longer looks: The effects of situation and gender. *The Quarterly Journal of Experimental Psychology, 63,* 1858-1871.

Macdonald, R. G., & Tatler, B. W. (2013a). Do as eye say: Gaze cueing and language in a real-world social interaction. *Journal of Vision, 13,* 1-12.

Macdonald, R. G., & Tatler, B. W. (2013b). The effect of social roles on gaze cue utilisation in a real-world collaboration. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.). *Proceedings of the 35th Annual Conference of the Cognitive Science Society.* (pp. 942-947). Austin, TX: Cognitive Science Society.

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision, 9,* 1-13.

Markus, H. (1977). The effect of mere presence on social facilitation: An unobtrusive test. *Journal of Experimental Social Psychology, 14,* 389-397.

McMains, S. A., & Somers, D. C. (2004). Multiple Spotlights of Attentional Selection in Human Visual Cortex. *Neuron, 42,* 677-686.

McMains, S. A., & Somers, D. C. (2005). Processing Efficiency of Divided Spatial Attention Mechanisms in Human Visual Cortex. *The Journal of Neuroscience, 25,* 9444-9448.

Müller, M. M., Malinowski, P., Gruber, T., & Hillyard, S. A. (2003). Sustained division of the attentional spotlight. *Nature, 424,* 309-312.

Neisser, U. (1967). *Cognitive Psychology.* New York: Appleton, Century, Crofts.

Nummenmaa, L., & Hietanen, J. K. (2006). Gaze distractors influence saccadic curvature: Evidence for the role of the oculomotor system in gaze-cued orienting. *Vision Research, 46,* 3674-3680.

Pashler, H. (1988). Familiarity and visual change direction. *Perception & Psychophysics, 44,* 369-378.

Pelphrey, K. A., Morris, J. P., & McCarthy, G. (2004). Grasping the Intentions of Others: The Perceived Intentionality of an Action Influences Activity in the Superior Temporal Sulcus during Social Perception. *Journal of Cognitive Neuroscience, 16,* 1706-1716.

Perrett, D. I., Hietanen, M. W., Oram, M. W., Benson, P. J., & Rolls, E. T. (1992). Organization and Functions of Cells Responsive to Faces in the Temporal Cortex [and Discussion]. *Philosophical Transactions of The Royal Society: B, 335,* 23-30.

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioural research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88,* 879-903.

Posner, M. I. (1980). Orienting of Attention. *Quarterly Journal of Experimental Psychology*, *32*, 3-25.

Posner, M. I., Synder, C., & Davidson, B. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109,* 160-174.

Pylyshyn, Z. W., Burkell, J., Fisher, B., Sears, C., Schmidt, W., & Trick, L. (1994). Multiple Parallel Access in Visual Attention. *Canadian Journal of Experimental Psychology, 48,* 260-283.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision, 3,* 179-197.

Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye movements in iconic visual search. *Vision Research, 42,* 1447-1463.

Rensink, R. A. (2000). Visual search for change: a probe into the nature of attentional processing. *Visual Cognition, 7,* 345-376.

Ricciardelli, P., Betta, E., Pruner, S., & Turatto, M. (2009). Is there a direct link between gaze perception and joint attention behaviours? Effects of gaze contrast polarity on oculomotor behaviour. *Experimental Brain Research, 194,* 347-357.

Ricciardelli, P., Bricolo, E., Aglioti, S. M. & Chelazzi, L. (2002). My eyes want to look where your eyes are looking: Exploring the tendency to imitate another individual's gaze. *Neuroreport, 13,* 2259-2264.

Ricciardelli, P., Carcagno, S., Vallar, G., & Bricolo, E. (2013). Is gaze following purely reflexive or goal-directed instead? Revisiting the automaticity of orienting attention by gaze cues. *Experimental Brain Research, 224,* 93-106.

Richardson, D. C., Hoover, M. A., & Ghane, A. (2008). Joint Perception: Gaze and the Presence of Others. *Proceeds of the 30th Annual Conference of the Cognitive Science Society.* (pp. 309-314). Washington DC, USA: Cognitive Science Society.

Richardson, D. C., Street, C. N. H., Tan, J. Y. M., Kirkham, N. Z., Hoover, M. A., & Cavanaugh, A. G. (2012). Joint perception: gaze and social context. *Frontiers in Human Neuroscience, 6,* 194. doi: 10.3389/fnhum.2012.00194

Risko, E. F., Laidlaw, K. E., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with Real vs. Reel stimuli: Toward an empirical approach to concerns about ecological validity. *Frontiers in Human Neuroscience, 6,* 143.

Ristic, J., & Kingstone, A. (2005). Taking control of reflexive social attention. *Cognition, 94,* B55-B65.

Sagiv, N., & Bentin, S. (2001). Structural Encoding of Human and Schematic Faces: Holistic and Part-Based Processes. *Journal of Cognitive Neuroscience, 13,* 937-951.

Schilling, H. E. H., Rayner, K., & Chumbley, J. L. (1983). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory and Cognition, 26,* 1270-1281.

Scholl, B. J., Pylyshyn, Z. W., & Feldmen, J. (2001). What is a visual object? Evidence from target merging in multiple object tracking. *Cognition, 80,* 159-177.

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *The Quarterly Journal of Experimental Psychology, 62,* 1904-1914.

Schneider, D., Nott, Z. E., & Dux, P. E. (2014). Task instructions and implicit theory of mind. *Cognition*, *133,* 43-47.

Schrammel, F., Pannasch, S., Graupner, S. T., Mojzisch, A., & Velichkovsky, B. M. (2009). Virtual friend or threat? The effects of facial expression and gaze interaction on psychophysiological responses and emotional experience. *Psychophysiology, 46,* 922-931.

Schultz, D. P. (1969). The human subject in psychological research. *Psychological Bulletin, 72,* 214-228.

Schwartz, S., Vuilleumier, P., Hutton, C., Maravita, A., Dolan, R. J., & Driver, J. (2005). Attentional Load and Sensory Competition in Human Vision: Modulation of fMRI Responses by Load at Fixation during Task-irrelevant Stimulation in the Peripheral Visual Field. *Cerebral Cortex, 15,* 770-786.

Sebanz, N. Bekkering, H., Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences, 10,* 70-76.

Shepherd, S. V. (2010). Following gaze: gaze-following behaviour as a window into social cognition. *Frontiers in Integrative Neuroscience,4,*1-13.

Shulman, G. L., Remington, R. W., & McLean, J. P. (1979). Moving attention through visual space. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 522-526.

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in Cognitive Sciences, 4,* 147-155.

Singer, T., Kiebel, S. J., Winston, J. S., Dolan, R. J., & Frith, C. D. (2004). Brain responses to the acquired moral status of faces. *Neuron, 41,* 653-662.

Smart, R. G. (1966). Subject selection bias in psychological research. *Canadian Psychologist/Psychologie Canadienne, 7a,* 115-121.

Smilek, D., Dixon, M. J. & Merikle, P. M. (2006). Revisiting the category effect: The influence of meaning and search strategy on the efficiency of visual search. *Brain Research, 1080,* 73-90.

Solman, G. J. F., Cheyne, J. A., & Smilek, D. (2011). Memory load affects visual search processes without influencing search efficiency. *Vision Research, 51,* 1185-1191.

Sorce, J. F., Emde, R. N., Campos, J. J., & Klinnert, M. D. (1985). Maternal emotional signalling: Its effect on the visual cliff behaviour of 1-year-olds. *Developmental Psychology, 21,* 195-200.

Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real world scenes. *Journal of Vision, 14,* 1-21.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences*, *13,* 403-409.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7,* 1-17.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*, 1-23.

Tatler, B. W., Kirtley, C., Macdonald, G., Mitchell, K. M. A., & Savage, S. (2013). The Active Eye: Perspectives on Eye Movement Research. In M. Horsley, M. Eliot, B. A. Knight, & R. Reilly (Eds.). *Current Trends in Eye Tracking Research* (pp.1-16). London, UK: Springer International Publishing.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research, 2,* 1-18.

Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition*, *17*, 1029-1054.

Tatler, B. W., Wade, N. J., Kwan, H., Findlay, J. M., & Velichkovsky, B. M. (2010). Yarbus, eye movements and vision. *i-Perception, 1,* 7-27.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics*, *49,* 83-90.

Tong, F. (2004). Splitting the Spotlight of Visual Attention. *Neuron, 42,* 524-526.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97-136.

Treisman, M., & Gellade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97-136.

Triplett, N. (1898). Dynamogenic Factors in Pacemaking and Competition. *The American Journal of Psychology, 9,* 507-533.

Tsal, Y. (1983). Movements of attention across the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 523-530.

Umiltà, C. (1988). Orienting of attention. In F. Boller, & J. Grafman (Eds.). *Handbook of Neuropsychology, Vol. 1.* (pp. 175- 193). Amsterdam: Elsevier.

VanRullen, R., Carlsion, T., & Cavanagh, P. (2007). The blinking spotlight of attention. *Proceedings of the National Academy of Sciences of the United States of America, 104,* 19204-19209.

Viviani, P., & Swensson, R. G. (1982). Saccadic eye movements to peripherally discriminated visual targets. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 113-126.

Vō, M. L. -H., & Schneider, W. X. (2010). A glimpse is not a glimpse: Differential processing of flashed scene previews leads to differential target search benefits. *Visual Cognition, 18,* 171-200.

Willis, J., & Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychological Science, 17,* 592-598.

Wilmmer, H., & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13,* 103-128.

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202-238.

Wolfe, J. M., Butcher, S. J., Lee, C., & Hyle, M. (2003). Changing your mind: On the contributions of top-down and bottom-up guidance in visual search for feature singletons. *Journal of Experimental Psychology: Human Perception and Performance, 29,* 483-502.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research, 44,* 1411-1426.

Woodman, G. F., Vogel, E. K., & Luck, S. J. (2001). Visual search is slowed when visuospatial working memory is occupied. *Psychonomic Bulletin and Review, 11,* 269-274.

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research, 49,* 2095-2103.

Yantis, S. & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 601-621.

Yantis, S. & Jonides, J. (1990). Abrupt Visual Onsets and Selective Attention: Voluntary Versus Automatic Allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 121-134.

Yarbus, A. L. (1967). *Eye Movements and Vision.* New York: Plenum Press.

Zajonc, R. B. (1965). Social facilitation. *Science, 149,* 269-274.

Zwickel, J., & Võ, M. L. -H. (2010). How the presence of persons biases eye movements. *Psychonomic Bulletin and Review, 17,* 257-262.